Metastable behavior in Markov processes with internal states

Jay Newby* Jon Chapman*

April 26, 2013

Abstract

A perturbation framework, called the quasi-stationary analysis (QSA), is developed to analyze metastable behavior in stochastic processes with random internal and external states. The QSA is illustrated with a model of gene expression that displays bistable switching. In this model, the external state represents the number of protein molecules produced by a hypothetical gene. Once produced, a protein is eventually degraded. The internal state represents the activated or unactivated state of the gene; in the activated state the gene produces protein more rapidly than the unactivated state. The gene is activated by a dimer of the protein it produces so that the activation rate depends on the current protein level. This is a well studied model, and several model reductions and diffusion approximation methods are available to analyze its behavior. However, it is unclear if these methods accurately approximate long-time metastable behavior (i.e., mean switching time between metastable states of the bistable system). Diffusion approximations are generally known to fail in this regard. It is shown that a diffusion approximation based on a quasi-steady-state reduction (stochastic averaging), which averages the internal state out and reduces the process to a continuous Markov process for the external state, provides unreliable accuracy. On the other hand, the QSA approximation is consistently accurate.

1 Introduction

A common feature found in many stochastic models of biological processes is a distinction between internal and external states [36]. There are numerous examples of such Markov processes used as models for biological phenomena [2, 10, 17, 27, 31]. Examples of an internal state include the number of open ion channels in the membrane of a neuron that affect its membrane voltage [16] and the on/off state of a gene that affects its protein production rate [28]. The distinction between internal and external states should not be confused with the concept of 'intrinsic' and 'extrinsic' noise (see Ref. [35] for an example related

*Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford, OX1 3LB, UK

to gene expression). A system with internal degrees of freedom is a classical idea in physics and applied mathematics, and the extension of this concept to Markov processes with internal states is well known in the literature [14,18,20].

The canonical example is the velocity jump process, which has a continuous external state, usually representing the physical location of a particle or some other observable quantity, that evolves deterministically with a velocity that depends on the current value of the internal state. The internal state evolves randomly according to a discrete jump Markov process. The simplest velocity jump process is called the telegraph process. The external state $X(t)$ evolves according to the stochastic differential equation

$$X(t) = X(0) + \int_0^t v(S(\tau))d\tau, \tag{1.1}$$

where the velocity $v(-1) = -v_-$ and $v(1) = v_+$. The internal state $S(t)$ can take the value ± 1 at any given time and is governed by the discrete Markov process

$$S(t) = S(0) + Y_1(\int_0^t \alpha\chi(S(\tau))d\tau) - Y_2(\int_0^t \beta(1-\chi(S(\tau)))d\tau), \tag{1.2}$$

where $Y_{1,2}(t)$ are independent, unit Poisson processes, and the indicator function $\chi(s) = 1$ if $s = 1$ and is zero otherwise. The transition rates (or intensities) α and β determine how quickly the particle switches velocity. Velocity jump processes have been used to model molecular motor transport [29], dynamic microtubule polymerization [2], and chemotaxis [31].

The Markov process becomes nonlinear when internal state transitions depend on the external state. For example, the process $(X(t), S(t))$ is nonlinear if the transition rates depend on $X(t)$ (i.e., $\alpha \to \alpha(X(t))$ and $\beta \to \beta(X(t))$). Interestingly, nonlinearities can lead to metastable behavior under so-called weak noise conditions. For the above example, this occurs when the transition rates are large compared to v/L, where L is a characteristic space scale on which the process is observed. In the large transition rate limit, the process becomes deterministic and

$$\dot{x} = \langle v \rangle \equiv \left(\frac{\alpha(x)}{\alpha(x)+\beta(x)}\right) v_+ - \left(\frac{\beta(x)}{\alpha(x)+\beta(x)}\right) v_-. \tag{1.3}$$

Metastable behavior is important because it represents fluctuation-induced phenomena that cannot be predicted by the deterministic limit. The standard example of metastable behavior is Brownian motion in a double well potential. On short timescales the particle is most likely found near one of the two minima, and on long timescales the particle can transition over the energy barrier that separates the two wells. Other types of Markov processes, such as the velocity-jump process, can behave like diffusion in a double well potential. Here, we consider processes with an internal state that evolves according to a discrete jump Markov process. But unlike the example velocity jump process described above, we also consider noise in the external state.

Assume that there is a well-defined limit where both noise sources are removed and the process converges to a deterministic system describing the external state. Under weak noise conditions, the dynamics of the deterministic system strongly influence the dynamics of the stochastic process. In particular, we are interested in the case where the deterministic system has multiple stable solutions depending on the initial conditions. On short timescales, a trajectory of the stochastic process fluctuates about the deterministic trajectory that has the same initial conditions. However, metastable behavior in the stochastic process is not seen in the deterministic system because it depends on a small amount of noise present in the system to cause a transition from one of the stable deterministic solutions to another. Metastable transitions occur on a long timescale.

Metastable transitions by nonlinear Markov processes with an internal and external state are more difficult to analyze than diffusion in a potential well, and exact analytical solutions are rarely possible. Monte Carlo simulation does not work either because metastable behavior requires far to much processor time to be practical. It is therefore necessary to develop approximation methods. One way to approximate is to eliminate a noise source. To eliminate noise in the internal state, an adiabatic limit is taken where the transitions are infinitely rapid so that the internal state can be averaged out. Eliminating noise in the external state results in a velocity jump process where the external state evolves deterministically in between random jumps in the internal state. However, the timescale for metastable transitions is very sensitive to both the type of noise and the noise strength, and eliminating a noise source can lead to large errors.

Another way to reduce the complexity of the model is with a quasi-steady-state (QSS) reduction. This is very similar to the adiabatic limit, but uses a perturbation approach so that higher order terms can be included that account for noise in the internal state. The QSS reduction includes effects from both noise sources, but approximates the coupled (internal, external) process with a single continuous Markov process for the external state. This approximation reduces the problem to diffusion in a double potential well. The underlying assumption behind the approximation is that the random internal state is well approximated by a random variable chosen from its steady-state distribution conditioned on a fixed external state. Intuitively, the assumption is valid if the internal state fluctuates much more rapidly than the external state (i.e., the small noise conditions described above).

While the QSS reduction is a useful tool in many circumstances, it is not accurate for characterizing metastability. Recently, in a model of molecular motor transport, metastable behavior was observed in a velocity jump process [27]. Following this observation, alternative perturbation methods were developed to analyze the metastable behavior [16,28,30]. However, these methods only work when the external state evolves deterministically as in the example described above. In this paper, we further develop this theory to account for noise in the external state.

To describe metastable behavior, it is necessary to approximate both the effective potential for the process and various timescales for metastable transi-

tions. For 1D continuous Markov processes, quantities such as the potential are straightforward to define, and if a stationary solution exists it must have zero probability flux. Given its usefulness at describing the qualitative features, we would like to know if we can define an effective potential in general. For higher dimensional continuous Markov processes, the potential well is no longer well defined when the drift velocity field is not the gradient of a potential. This is closely related to detailed balance conditions and thermodynamic equilibrium. In higher dimensions, it is possible for the stationary density to exhibit a nonzero probability flux. Developing a systematic formalism to describe nonequilibrium stationary behavior is particularly relevant in biology. It turns out that an effective potential can still be defined through perturbation theory [32] and transition path theory [13, 25]. These tools can also be used to approximate the timescale associated with metastable transitions. The methods presented here fit within the perturbation framework.

The theory of large deviations [8,9,33] is the mathematical foundation for the techniques used to study metastable transitions (rare events). Here, we focus on perturbation-theory-based techniques [32], which we refer to as the quasi-stationary analysis (QSA). The QSA was developed to analyze the differential Kolmogorov equation [11], which describes the process through its probability density function. For a continuous Markov process, the QSA is well-developed for the Fokker–Planck equation [22–24, 26, 32, 34]. The QSA has also be applied to the Master equation to analyze certain birth-death processes [1,3–7,12,15,37]. (The Fokker–Planck and Master equation are instances of the more general differential Kolmogorov equation.) However, as far as we are aware, no one has applied these methods to study weak noise problems where QSS reduction (i.e., stochastic averaging) and the large system size limit are both necessary to reach the deterministic system. Furthermore, large deviation theory does not provide a means of explicitly calculating the preexponential factor (see Section 4.1), which is a significant part of the leading order transition time and stationary density approximations.

There are several advantages to the QSA over diffusion-approximation and model-reduction techniques. First, if the process includes a discrete state, the QSA provides an approximation that accounts for all moments of the jump propagator, whereas the diffusion approximations include only the first two moments (e.g., a diffusion approximation of a discrete jump Markov process by truncation of a Kramers–Moyal (KM) expansion). Second, the approximation provides a uniformly accurate approximation of the stationary probability density function. Third, physically meaningful quantities, such as the effective potential and metastable transition rates, can be generalized to processes that do not assume detailed balance. Finally, the QSA can be applied to high dimensional (in the external state) problems displaying nonequilibrium stationary behavior.

In this paper, we develop the QSA for a general class of Markov processes that have a discrete internal state, and we illustrate the analysis using a simple model problem. The model problem is ideal because it is possible to derive a QSS diffusion approximation and several different model reductions, allowing us to (i) develop the general QSA procedure for different types of Markov processes

(ii) compare the resulting approximations with the QSS diffusion approximation. In particular, we are interested in approximating two quantities: the timescales for metastable transitions and the effective potential. The analysis of the model problem should inform our understanding about when reduction techniques, such as a diffusion approximation, fail to approximate these two quantities and why. Previous work has shown that diffusion approximations lead to errors in both the adiabatic [12, 38] and quasi-deterministic limits [28]. But what happens when both noise sources are present? When is one noise source more significant than the other? If the QSS reduction fails, why does it fail? Is it due to large deviation errors like the system-size expansion, or is it because the QSS assumption is invalid? Does it fail for the same reasons in each limit?

The paper is organized as follows. First, in Section 2 we introduce the model problem along with various approximations and model reductions. In Section 3 we describe the dynamics of the deterministic limit and explore the two limits where noise in the either the internal or external state is removed. The QSA is presented in Section 4, and results are shown in Section 5.

2 Model problem

Consider the following as an example of a discrete Markov process with an internal state. A population of proteins changes stochastically according to a jump Markov process, where the protein production rate depends on the internal state of the system. The hypothetical gene responsible for producing the protein has a promotor that can be turned on or off. When the promotor is on, protein is produced at higher rate than when it is off. For simplicity all parameters are presented in nondimensional form. The following state diagram, where N_n is the state where n proteins are present in the system, represents the external state transitions:

$$N_0 \underset{\delta}{\overset{\tau(S(t))}{\rightleftarrows}} N_1 \underset{2\delta}{\overset{\tau(S(t))}{\rightleftarrows}} N_2 \cdots \underset{n\delta}{\overset{\tau(S(t))}{\rightleftarrows}} N_n \underset{(n+1)\delta}{\overset{\tau(S(t))}{\rightleftarrows}} \cdots, \tag{2.1}$$

where we set $\delta = 1$ and $S(t)$ is the two state stochastic process governing the on/off state of the promotor; $S(t) = 1$ when the promoter is on and $S(t) = 0$ when it is off. The production rate is a function of the promotor state, with

$$\tau(0) = \sigma\alpha_{\mathrm{e}}, \quad \tau(1) = \alpha_{\mathrm{e}}. \tag{2.2}$$

The nondimensional parameter σ controls how much spontaneous protein production occurs when the promotor is off, and we assume $0 < \sigma < 1$ so that protein production is higher when the promotor is on. To get nonlinear phenomena, the internal state transitions must depend on the external state. A simple model of the promotor fluctuations is given by

$$(\text{off}) \underset{\alpha_{\mathrm{i}}\beta}{\overset{\alpha_{\mathrm{i}}N(t)^2/\alpha_{\mathrm{e}}^2}{\rightleftarrows}} (\text{on}), \tag{2.3}$$

where $N(t)$ is the number of protein copies. This is a model of a promotor that can be turned on by a dimer of the protein product [17].

To fully model the dimerization reaction, an additional state variable is required which expands the state space of the process. Under the assumption that there are much fewer dimers than monomers, a quasi-steady-state approximation can be made [17, 19]. Of course, this is a model reduction approximation, and we do not fail to recognize that this is a paper about the validity model reduction techniques. However, because we are interested in a simple model we can use to explore various mathematical issues, we accept (2.3) as the basic model of promotor switching, though we make no claims about its accuracy in approximating the dimerization reaction.

The transitions between the two promotor states are assumed to be fast by specifying that $1/\alpha_{\rm i} \ll 1$ is a small parameter. Thus, the full Markov process for the combined external/internal state of the system is $\mathcal{M}(t) = (N(t), S(t))$. Note that neither process is Markovian by itself. The Chapman-Kolmogorov (CK) equation governing the probability density function $\mathrm{p}_j(n,t) \equiv \mathrm{p}(j,n,t \mid j_0, n_0, t_0)$ for $\mathcal{M}(t)$ is

$$\frac{\partial}{\partial t}\mathrm{p}_0(n,t) = \left[(\mathbb{E}^+ - 1)n + \alpha_{\rm e}\sigma(\mathbb{E}^- - 1)\right]\mathrm{p}_0 + \alpha_{\rm i}\left(-\frac{n^2}{\alpha_{\rm e}^2}\mathrm{p}_0 + \beta \mathrm{p}_1\right) \tag{2.4a}$$

$$\frac{\partial}{\partial t}\mathrm{p}_1(n,t) = \left[(\mathbb{E}^+ - 1)n + \alpha_{\rm e}(\mathbb{E}^- - 1)\right]\mathrm{p}_1 + \alpha_{\rm i}\left(\frac{n^2}{\alpha_{\rm e}^2}\mathrm{p}_0 - \beta \mathrm{p}_1\right), \tag{2.4b}$$

where the jump operators $\mathbb{E}^{\pm}$ are defined by

$$\mathbb{E}^{\pm} f(n) = f(n \pm 1). \tag{2.5}$$

It is convenient to introduce the following notation. We can rewrite the probability density function as

$$\begin{aligned}\mathrm{Prob}[S(t) = s, X(t) \in (x, x+dx)] = \mathrm{Prob}[S(t) = s | X(t) \in (x, x+dx)] \\ \times \mathrm{Prob}[X(t) \in (x, x+dx)].\end{aligned} \tag{2.6}$$

Define the conditional internal state distribution to be

$$\mathrm{w}_s(x,t) \equiv \mathrm{Prob}[s = S(t) | X(t) \in (x, x+dx)], \tag{2.7}$$

and the marginal external state density to be

$$u(x,t) \equiv \mathrm{Prob}[X(t) \in (x, x+dx)]. \tag{2.8}$$

Introduce vector notation for the probability density with

$$\mathbf{p}(x,t) \equiv (\mathrm{p}_0(x,t), \mathrm{p}_1(x,t))^T. \tag{2.9}$$

We write the stationary versions of (2.7)-(2.9) as

$$\lim_{t\to\infty} \mathbf{p}(x,t) = \hat{\mathbf{p}}(x) = \hat{\mathbf{w}}(x)\hat{u}(x), \tag{2.10}$$

where $\hat{\mathbf{w}}(x)$ is a 2-vector and

$$\sum_s \hat{w}_s(x) = 1. \tag{2.11}$$

2.1 Semi-continuous process

A closely-related process assumes an external state that evolves according to a continuous Markov process instead of a discrete birth-death process as in the previous section. Consider the Langevin equation

$$dX(t) = -v(S(t))dt + \sqrt{b(S(t))}dW(t). \tag{2.12}$$

where $S(t)$ is again the internal state. While this process can be quite general, in order to make the connection to the discrete process (2.1) we define it as the diffusion approximation via a truncated KM expansion. The mean number of proteins when the promotor is on is α_{e}. When α_{e} is large we can rescale to a continuous variable $X(t) = N(t)/\alpha_{\mathrm{e}}$. The internal state $S(t)$ then evolves according to

$$(\text{off}) \underset{\alpha_{\mathrm{i}}\beta}{\overset{\alpha_{\mathrm{i}}X(t)^2}{\rightleftarrows}} (\text{on}), \tag{2.13}$$

It is straight forward to show that the drift in each state is

$$v(0) = \sigma - X, \quad v(1) = 1 - X, \tag{2.14}$$

and the diffusivities are

$$b(0) = \frac{1}{\alpha_{\mathrm{e}}}(\sigma + X), \quad b(1) = \frac{1}{\alpha_{\mathrm{e}}}(1 + X). \tag{2.15}$$

The full Markov process is then $\mathcal{C}(t) = (X(t), S(t))$, and its corresponding differential Kolmogorov equation is

$$\frac{\partial}{\partial t}\mathrm{p}_0(x,t) = -\frac{\partial}{\partial x}[(\sigma - x)\mathrm{p}_0] + \frac{1}{2\alpha_{\mathrm{e}}}\frac{\partial^2}{\partial x^2}[(\sigma + x)\mathrm{p}_0] - \alpha_{\mathrm{i}}\left(x^2\mathrm{p}_0 - \beta\mathrm{p}_1\right) \tag{2.16a}$$

$$\frac{\partial}{\partial t}\mathrm{p}_1(x,t) = -\frac{\partial}{\partial x}[(1 - x)\mathrm{p}_1] + \frac{1}{2\alpha_{\mathrm{e}}}\frac{\partial^2}{\partial x^2}[(1 + x)\mathrm{p}_1] + \alpha_{\mathrm{i}}\left(x^2\mathrm{p}_0 - \beta\mathrm{p}_1\right). \tag{2.16b}$$

2.2 Diffusion approximation

There is a standard asymptotic method available to analyze both processes introduced above, based on the same principle as that of a quasi-steady-state (QSS) reduction. If the system transitions between the internal states much more quickly than it transitions between its external states, then the internal states are approximately given by their steady state distribution conditioned on a fixed external state. The combined process $(X(t), S(t))$, can be reduced using a projection method to eliminate (average out) the internal states $S(t)$ to obtain a Markov process that approximates the external state $X(t)$. In other words, although $X(t)$ is not strictly Markovian, it may be approximately Markovian.

A projection method results in a scalar Fokker–Planck equation for the marginal external-state probability density function

$$u(x,t) = \sum_{s=0}^{1} \mathrm{p}_s(x,t). \tag{2.17}$$

For a general discussion of the QSS projection method see [11]. For brevity we only quote the result here; for further details of the QSS reduction of the model problem considered here, we refer the reader to [17]. The result is

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}(\langle \mathbf{v}(x)\rangle\, u) + \frac{1}{\alpha_{\mathrm{i}}}\frac{\partial^2}{\partial x^2}\left(\langle \mathbf{b}(x)\rangle\, u\right) + \frac{1}{\alpha_{\mathrm{i}}}\frac{\partial}{\partial x}\left(D(x)\frac{\partial u}{\partial x}\right), \tag{2.18}$$

where

$$\langle \mathbf{v}(x)\rangle = \frac{\beta(\sigma - x)}{\beta + x^2} + \frac{x^2(1-x)}{\beta + x^2}, \tag{2.19}$$

$$\langle \mathbf{b}(x)\rangle = \frac{\alpha_{\mathrm{i}}}{2\alpha_{\mathrm{e}}}\left(\frac{\beta(\sigma + x)}{\beta + x^2} + \frac{x^2(1+x)}{\beta + x^2}\right), \tag{2.20}$$

$$D(x) = \frac{\beta((\sigma - x) - \langle \mathbf{v}(x)\rangle)(\sigma - x)}{(\beta + x^2)^2} + \frac{x^2((1-x) - \langle \mathbf{v}(x)\rangle)(1-x)}{(\beta + x^2)^2}. \tag{2.21}$$

Protein fluctuations are captured by $\langle \mathbf{b}(x)\rangle$ and promotor fluctuations by $D(x)$.

We can rewrite the Fokker–Planck equation (2.18) in conservation form, with

$$\frac{\partial u}{\partial t} = -\frac{\partial J}{\partial x}, \quad J(x,t) = a(x)u(x,t) - B(x)\frac{\partial u}{\partial x}, \tag{2.22}$$

where

$$a(x) = \langle \mathbf{v}(x)\rangle - \frac{\langle \mathbf{b}(x)\rangle'}{\alpha_{\mathrm{i}}}, \quad B(x) = \frac{1}{\alpha_{\mathrm{i}}}\left(\langle \mathbf{b}(x)\rangle + D(x)\right). \tag{2.23}$$

The mean first passage time (MFPT) for any given starting position x_0 to reach $x_\pm$ is given exactly by

$$T = \int_{x_0}^{x_\pm}\int_0^y \frac{e^{\Theta(z)-\Theta(y)}}{B(y)}\,dz\,dy, \quad \Theta(x) = \int_0^x \frac{a(y)}{B(y)}\,dy. \tag{2.24}$$

To make comparisons to other approximations, we make the following definitions. Let

$$\Phi(x) = \int_{x_\pm}^{x} \frac{-\langle \mathbf{v}(y)\rangle}{\langle \mathbf{b}(y)\rangle + D(y)}\,dy, \quad \Psi(x) = \int_{x_\pm}^{x} \frac{\frac{d}{dy}\langle \mathbf{b}(y)\rangle}{\langle \mathbf{b}(y)\rangle + D(y)}\,dy. \tag{2.25}$$

The MFPT is then approximated by $T_\pm \equiv T(x_\pm) \sim 1/\lambda_\pm$, where

$$\lambda_\pm = \left(\frac{B(x_*)}{\pi}\sqrt{|\Phi''(x_*)|\,\Phi''(x_\pm)}\right) e^{-\Psi(x_*)} \exp\left[-\frac{1}{\epsilon}\Phi(x_*)\right]. \tag{2.26}$$

3 Limiting processes

In the previous section, we define the model problem. We call the full version presented in Section 2 the discrete process, which makes no assumption about the size of the parameters α_{i} or α_{e}. The semi-continuous process from Section 2.1 is an approximation of the discrete process where $x = n/\alpha_{\mathrm{e}}$ is treated as a continuous variable and we assume that $\alpha_{\mathrm{e}} \gg 1$ is a large parameter; no assumption is made about the size of α_{i}. If we assume that α_{i} is also a large parameter then further reduction is possible using a diffusion approximation, call it the QSS process, which is presented in Section 2.2. All three versions contain terms that depend on α_{i} and α_{e}, and if these parameters are assumed to be large, all three should account for contributions of noise in the external and internal state.

Of course, further model reduction is possible by removing one source of noise: either $\alpha_{\mathrm{i}} \to \infty$ or $\alpha_{\mathrm{e}} \to \infty$. The former is known in the literature as the adiabatic limit (see [17,38]), and the later we call the quasi-deterministic (QD) limit. If both limits are taken, a well defined deterministic dynamical system is obtained. Note that all three versions of the model problem—the discrete, semi-continuous, and QSS processes—converge to the same deterministic limit (3.1). In the rest of this section we explore each limit in turn.

3.1 Deterministic limit $\alpha_{\mathrm{i}} \to \infty$ and $\alpha_{\mathrm{e}} \to \infty$

If we take the limit $\alpha_{\mathrm{e}} \to \infty$ and $\alpha_{\mathrm{i}} \to \infty$, the resulting deterministic dynamics is

$$\dot{x} = \frac{\beta(\sigma - x) + x^2(1 - x)}{\beta + x^2}. \tag{3.1}$$

Assuming that $\sigma \ll 1$, the system is described as follows. For $\beta_- < \beta < \beta_+$, where

$$\beta_- \sim 4\sigma + O(\sigma^2), \quad \beta_+ \sim \frac{1}{4} + \frac{\sigma}{2} + O(\sigma^2) \tag{3.2}$$

the system is bistable, with an unstable fixed point at

$$x_* \sim \frac{1}{2}(1 - \sqrt{1 - 4\beta}) + O(\sigma) \tag{3.3}$$

and two stable fixed points at

$$x_- \sim \sigma + O(\sigma^2), \quad x_+ \sim \frac{1}{2}(1 + \sqrt{1 - 4\beta}) + O(\sigma) \tag{3.4}$$

This is the regime of interest as we wish to characterize the transition times between the two stable fixed points when the system is stochastic with weak fluctuations.

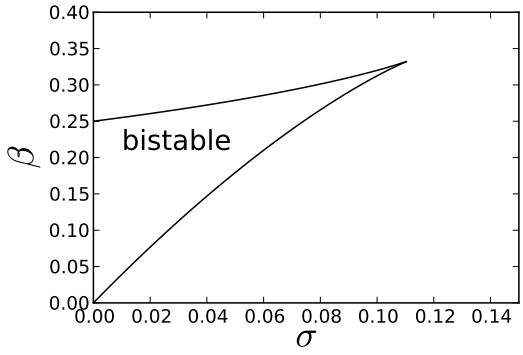

Figure 1: Bifurcation diagram for the deterministic dynamics.

3.2 Quasi-deterministic limit $\alpha_e \to \infty$

A velocity jump process, as described in the Introduction, can be obtained by taking the limit $\alpha_e \to \infty$ in either the discrete or semi-continuous process (both processes converge in this limit). This limit is discussed in [17] and later a metastable analysis was introduced in [28]. In this limit, the differential Kolmogorov equation for both the discrete and semi-continuous processes converges to

$$\frac{\partial}{\partial t}\mathrm{p}_0(x,t) = -\frac{\partial}{\partial x}[(\sigma - x)\mathrm{p}_0] - \alpha_i\left(x^2\mathrm{p}_0 - \beta\mathrm{p}_1\right) \tag{3.5a}$$

$$\frac{\partial}{\partial t}\mathrm{p}_1(x,t) = -\frac{\partial}{\partial x}[(1 - x)\mathrm{p}_1] + \alpha_i\left(x^2\mathrm{p}_0 - \beta\mathrm{p}_1\right). \tag{3.5b}$$

It is worth pointing out that the QSS approximation converges to a very different process in this limit; the differential Kolmogorov (in this case Fokker–Planck) equation becomes

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}(\langle \mathbf{v}(x)\rangle\, u) + \frac{1}{\alpha_i}\frac{\partial}{\partial x}\left(D(x)\frac{\partial u}{\partial x}\right), \tag{3.6}$$

where $\langle \mathbf{v}(x)\rangle$ and $D(x)$ are given by (2.19) and (2.21), respectively. Moreover, one can derive (3.6) as a QSS approximation of (3.5) using the same projection methods as were used to derive (2.18).

3.3 Adiabatic limit $\alpha_i \to \infty$

The reduction procedure is based on a projection method very similar to the QSS reduction in Section 2.2. We leave the details to Appendix B and state the result. The limiting master equation is

$$\frac{du}{dt} = \alpha_e \mathbb{K} u, \tag{3.7}$$

where

$$u(n,t) \equiv \sum_{s=0}^{1} \mathrm{p}_s(n,t), \tag{3.8}$$

and

$$\mathbb{K} \equiv (\mathbb{E}^+ - 1)\frac{n}{\alpha_\mathrm{e}} + (\mathbb{E}^- - 1) f(\frac{n}{\alpha_\mathrm{e}}), \quad f(x) = \frac{\sigma\beta + x^2}{\beta + x^2}. \tag{3.9}$$

Note that at deterministic fixed points, x_c, $c = *, \pm$, $f(x_c) = x_c$.

4 Long-time asymptotic approximation of the probability density function

So far we have introduced the model problem, the discrete process, along with two approximations, the semi-continuous and QSS processes, which account for both noise sources. We also presented limiting processes that result when one of the two noise sources is removed, resulting in either the adiabatic or QD limiting processes. We now present a systematic perturbation method to analyze metastable, or long-time, behavior of the discrete and semi-continuous processes.

Suppose we have a CK equation of the form

$$\frac{\partial}{\partial t}\mathrm{p}(x,s,t) = -\mathcal{L}_\epsilon \mathrm{p}, \tag{4.1}$$

where $\mathcal{L}_\epsilon$ is a linear operator acting on state variables (x,s). The state variables are split so that $x \in \Omega \subset \mathbb{R}$ represents the external state and $s \in \mathbb{N}$ represents the internal state. Note that one can easily generalize this theory to the case $x \in \Omega \subset \mathbb{R}^N$ (see [28]). For illustration, take $\mathcal{L}_\epsilon$ to have the form

$$\begin{aligned}\mathcal{L}_\epsilon \mathrm{p} \equiv \frac{\partial}{\partial x}(v(x,s)\mathrm{p}) - \epsilon\frac{\partial^2}{\partial x^2}(b(x,s)\mathrm{p}) \\ - \frac{1}{\epsilon}\sum_{s'}\left(a(x,s|s')\mathrm{p}(x,s',t) - a(x,s'|s)\mathrm{p}(x,s,t)\right),\end{aligned} \tag{4.2}$$

where $\epsilon \ll 1$ is a small parameter (we will define ϵ for the model problem later in this section). The CK can also be written in conservation form with

$$\frac{\partial}{\partial t}u(x,t) = -\frac{\partial}{\partial x}J(x,t), \tag{4.3}$$

where

$$u(x,t) \equiv \sum_s \mathrm{p}(x,s,t), \quad J(x,t) \equiv \sum_s \left(v(x,s)\mathrm{p}(x,s,t) - \epsilon\frac{\partial}{\partial x}b(x,s)\mathrm{p}(x,s,t)\right).$$

(4.4)

Assume that $\mathcal{L}_\epsilon$ has a complete set of eigenfunctions, $\{\phi_j(x,s)\}$ and adjoint eigenfunctions $\{\xi_j(x,s)\}$. If the initial condition is $\mathrm{p}(x,s,0) = \delta(x-x_0)\delta(s-s_0)$, the solution can be written

$$\mathrm{p}(x,s,t) = \sum_{j=0}^{\infty} \xi_j(x_0,s_0)\phi_j(x,s)e^{-\lambda_j t}, \tag{4.5}$$

where we assume that all of the eigenvalues, λ_j, are nonnegative. Since we are interested in metastable behavior, assume that in the limit $\epsilon \to 0$, a deterministic, bistable process is reached for the external state x. Label the two stable fixed points $x_\pm$ and the unstable fixed point x_* and assume $x_- < x_* < x_+$.

The random process will look very different if the external state starts at $x_0 < x_*$ or $x_0 > x_*$. For the sake of illustration assume that $x_0 = x_-$. On intermediate time scales, the solution will converge to a stationary density around x_- that, figuratively speaking, does not see the other stable fixed point—or said another way, the solution does not see beyond x_*. Slowly, over a long timescale, the solution converges to the full stationary density as probability slowly leaks out past x_* toward x_+. The timescale for this long-time convergence is exponentially large (i.e., $O(e^{1/\epsilon})$). Since a stationary solution exists, the smallest, or principle, eigenvalue is $\lambda_0 = 0$, and the stationary density is the eigenfunction $\phi_0(x,s)$; that is, we normalize the principal eigenfunction so that $\langle \phi_0, 1\rangle = 1$ with respect to the inner product defined by

$$\langle f(x,s), g(x,s)\rangle \equiv \int_{x\in\Omega} \sum_s f(x,s)g(x,s)dx. \tag{4.6}$$

The separation of time scales in the problem can be exploited to approximate the solution. To understand how this works consider the process where a boundary condition is placed at x_* so that the process truly does not see beyond the unstable fixed point. We want to consider two different boundary conditions: reflecting and absorbing. To distinguish between each case, we write the principal eigenvalue and eigenfunction as $\lambda^{(a)}$, $\phi^{(a)}$ and $\lambda^{(r)}$, $\phi^{(r)}$ for absorbing and reflecting boundary conditions, respectively. If we place a reflecting boundary at x_* the principal eigenvalue $\lambda^{(r)} = 0$, but the eigenfunction $\phi^{(r)}$ is now restricted to $x \in \Omega_- = (-\infty, x_*)$ (or $x \in \Omega_+ = (x_*, \infty)$ if we instead assume that $x_0 > x_*$). We call this the quasi-stationary density.

Now suppose that an absorbing boundary is imposed at x_*. In this case, no stationary density exists, and the principal eigenvalue is perturbed by an exponentially small amount, that is, $\lambda^{(a)} = O(e^{-C/\epsilon})$, for some $C > 0$. The eigenfunction $\phi^{(a)}$ is also perturbed, but away from the boundary, $\phi^{(a)} \sim \phi^{(r)}$, which turns out to be straight forward to compute using a WKB approximation method. Thus, if we can calculate the eigenvalue and eigenfunction, we have an accurate approximation to the absorbing boundary problem with

$$\mathrm{p}(x,s,t) \sim \phi^{(a)}(x,s)e^{-\lambda^{(a)}t}, \quad t\lambda_1^{(a)} \gg 1, \tag{4.7}$$

or, since $\phi^{(a)} \sim \phi^{(r)}$,

$$\mathrm{p}(x,s,t) \sim \phi^{(r)}(x,s)e^{-\lambda^{(a)}t}, \quad t\lambda_1^{(a)} \gg 1, \quad \epsilon \ll 1. \tag{4.8}$$

We discuss how to approximate $\lambda^{(a)}$ later in this section.

This approximation can be repeated for the initial condition $x_0 > x_+$, and a different principle eigenvalue and quasi-stationary density are obtained, call the eigenvalues $\lambda_\pm^{(a)}$ and quasi-stationary densities $\phi_\pm^{(r)}(x,s)$. The full system, without any boundary condition imposed at x_*, can then be approximated by

$$\mathrm{p}(x,s,t) \sim \frac{1}{2}\begin{cases} q_-(t)\phi_-^{(r)}(x,s), & x < x_* \\ q_+(t)\phi_+^{(r)}(x,s), & x > x_* \end{cases}, \tag{4.9}$$

where $q_\pm(t)$ satisfy the system of ordinary differential equations

$$\frac{dq_-}{dt} = -\lambda_-^{(a)}q_- + \lambda_+^{(a)}q_+ \tag{4.10}$$

$$\frac{dq_+}{dt} = \lambda_-^{(a)}q_- - \lambda_+^{(a)}q_+, \tag{4.11}$$

with $q_-(0) = 1$ and $q_+(0) = 0$ if $x_0 < x_*$, or $q_-(0) = 0$ and $q_+(0) = 1$ if $x_0 > x_*$.

To obtain an approximation of $\lambda^{(a)}$, we use a spectral projection method that makes use of the adjoint operator $\mathcal{L}_\epsilon^*$. This method was first developed for scalar-value PDE eigenvalue problems [15,21] and later generalized to a vector-valued PDE eigenvalue problem [16,28,30]. The present treatment further generalizes the method. Consider the adjoint eigenfunctions $\{\xi_j\}$, $j = 0, 1, \cdots$, satisfying

$$\mathcal{L}_\epsilon^*\xi_j = \lambda_j\xi_j, \tag{4.12}$$

and take $\langle\phi_i, \xi_j\rangle = \delta_{i,j}$ so that the two sets of eigenfunctions are biorthogonal. We use the same notation to distinguish between the two boundary conditions for the adjoint eigenfunction. If the boundary is reflecting, the first adjoint eigenfunction is $\xi^{(r)} = 1$, and if the boundary is absorbing then $\xi^{(a)} \sim \xi^{(r)}$ away from the boundary, but develops a boundary layer at x_*. Using integration by parts we have

$$\left\langle \phi^{(r)}, \lambda^{(a)}\xi^{(a)} \right\rangle = \left\langle \phi^{(r)}, \mathcal{L}_\epsilon^*\xi^{(a)} \right\rangle = \left\langle \mathcal{L}_\epsilon\phi^{(r)}, \xi^{(a)} \right\rangle + J(\phi^{(r)}, \xi^{(a)}), \tag{4.13}$$

where the boundary contribution,

$$J(\phi^{(r)}, \xi^{(a)}) = \epsilon \sum_s b(x_*, s)\phi^{(r)}(x_*, s)\frac{d}{dx}\xi^{(a)}(x_*, s). \tag{4.14}$$

is nonzero because $\phi^{(r)}$ does not satisfy the absorbing boundary condition. Then, since $\mathcal{L}_\epsilon\phi^{(r)} = 0$, the principal eigenvalue is

$$\lambda^{(a)} = \frac{J(\phi^{(r)}, \xi^{(a)})}{\left\langle \phi^{(r)}, \xi^{(a)} \right\rangle}. \tag{4.15}$$

The above identity can be used to approximate the principal eigenvalue as follows. Since away from the boundary x_*, $\xi^{(a)} \sim \xi^{(r)} = 1$, we can make this substitution for the term in the denominator of (4.15) so that

$$\lambda^{(a)} \sim \frac{J(\phi^{(r)}, \xi^{(a)})}{\langle \phi^{(r)}, 1 \rangle}, \tag{4.16}$$

with exponentially small error. Notice that the denominator is then well approximated by the normalization factor for the eigenfunction $\phi^{(r)}$. We cannot make the same substitution for the term in the numerator, since (4.15) becomes a formula for $\lambda^{(r)} = 0$ instead of $\lambda^{(a)}$. Of course, in some sense zero is actually a very good approximation because the error is $O(e^{-C/\epsilon})$, but to capture the metastable behavior we need to capture the small exponential. Note that we could have just as well used $\phi^{(a)}$ and $\xi^{(r)}$ in (4.13) instead of $\phi^{(r)}$ and $\xi^{(a)}$. We choose the later because it simplifies the boundary layer analysis.

The recipe for approximating the solution requires approximations of the first eigenfunction and the first adjoint eigenfunction, where the latter satisfies the appropriate adjoint absorbing boundary condition. For the examples considered here, the eigenfunctions are treated as vector-valued functions of x. That is, the internal state is restricted to $s \in \{0, 1\}$ and we write $\phi_0(x, s) = [\boldsymbol{\phi}_0(x)]_s$ with $\boldsymbol{\phi}_0 : \mathbb{R} \mapsto \mathbb{R}^2$.

4.1 WKB approximation of the eigenfunction $\phi^{(r)}(x)$

Before proceeding with the perturbation analysis, it is convenient to make a few definitions. The CK equation for either the discrete process (2.4) or the continuous process (2.16) can be written in matrix form. Specifically, we are interested in the properties of the transition rate matrix for the internal state. Based on the transition rate diagram (2.13) the transition rate matrix is

$$A(x) \equiv \begin{bmatrix} -x^2 & \beta \\ x^2 & -\beta \end{bmatrix}. \tag{4.17}$$

Transition rate matrices like this are members of a family of matrices called $\mathbb{W}$-matrices. Notice that the columns sum to zero, which means that the matrix is singular and the vector,

$$\mathbf{1} \equiv \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \tag{4.18}$$

is the left eigenvector corresponding to a zero eigenvalue. For a transition rate matrix to be a $\mathbb{W}$-matrix, it must have negative diagonal elements, nonnegative off-diagonal elements, and it must be irreducible. One can show, using the *Perron-Frobenius theorem*, that the nullspace of a $\mathbb{W}$-matrix is one dimensional and that the right nullvector has strictly positive elements. We denote this vector as $\boldsymbol{\rho}$, where $A\boldsymbol{\rho} = 0$. Note that if the matrix A is a function of x then so is $\boldsymbol{\rho}$. For the matrix (4.17) we have that

$$\boldsymbol{\rho}(x) = \begin{bmatrix} \frac{\beta}{\beta + x^2} \\ \frac{x^2}{\beta + x^2} \end{bmatrix}, \tag{4.19}$$

where we have normalized $\boldsymbol{\rho}$ so that its elements sum to unity. For a simple two-state Markov process, where the external state is fixed, $\boldsymbol{\rho}$ is the steady state distribution.

The WKB analysis is carried out in the small variable ϵ defined as follows. Assume that the ratio between the two large parameters, $\alpha_{\rm e}$ and $\alpha_{\rm i}$, is fixed so that $\alpha_{\rm i} = \varphi\alpha_{\rm e}$ for fixed φ and define the small parameter $\epsilon = 1/\alpha_{\rm i} = 1/(\varphi\alpha_{\rm e})$. It is convenient to put the CK equation (2.4) or (2.16) in matrix form so that the eigenfunction (up to terms exponentially small in ϵ) satisfies

$$[A(x) + \Sigma_{\mathbb{d}}]\,\boldsymbol{\phi}_0(x) = 0, \tag{4.20}$$

where A, given by (4.17), is the transition rate matrix for the internal state and $\mathbb{d}$ is a vector of linear operators governing transitions between external states. For the discrete process we have

$$\mathbb{d}_{\rm disc} \equiv \frac{1}{\varphi}\begin{bmatrix}(\mathbb{e}^{\partial x}-1)x + \sigma(\mathbb{e}^{-\partial x}-1)\\ (\mathbb{e}^{\partial x}-1)x + (\mathbb{e}^{-\partial x}-1)\end{bmatrix}, \tag{4.21}$$

where the jump operators,

$$\mathbb{e}^{\pm\partial x} f(x) \equiv \sum_{n=0}^{\infty}\frac{(\pm\varphi\epsilon)^n}{n!}f^{(n)}(x) = f(x\pm\varphi\epsilon), \tag{4.22}$$

are defined in terms of a Taylor series expansion. For the semi-continuous process

$$\mathbb{d}_{\rm sc} \equiv -\epsilon\frac{d}{dx}\Sigma_{\mathbf{v}(x)} + \epsilon^2\frac{d^2}{dx^2}\Sigma_{\mathbf{b}(x)}, \tag{4.23}$$

where the drift and diffusivities are contained in

$$\mathbf{v}(x) \equiv \begin{bmatrix}\sigma - x\\ 1-x\end{bmatrix}, \quad \mathbf{b}(x) \equiv \begin{bmatrix}\frac{(\sigma+x)\varphi}{2}\\ \frac{(1+x)\varphi}{2}\end{bmatrix}. \tag{4.24}$$

We assume that the eigenfunction has the following WKB form

$$\boldsymbol{\phi}_0(x) \sim (\mathbf{r}_0(x) + \epsilon\mathbf{r}_1(x) + \cdots)\exp\left[-\frac{1}{\epsilon}\Phi(x)\right], \tag{4.25}$$

where Φ is a scalar functions and $\mathbf{r}_{0,1}$ are 2-vectors (with $\mathbf{r}_0$ positive). Substituting (4.25) into (4.20) and collecting leading order terms yields

$$O(1): \quad \left[A(x) + \Sigma_{\mathbf{h}(x,\Phi'(x))}\right]\mathbf{r}_0(x) = 0, \tag{4.26}$$

where, for the discrete process

$$\mathbf{h}_{\rm disc}(x,p) = \frac{1}{\varphi}(e^{\varphi p}-1)\left(\mathbf{v}(0) - xe^{-\varphi p}\right), \tag{4.27}$$

and for the semi-continuous process

$$\mathbf{h}_{\rm sc}(x,p) = p\mathbf{v}(x) + p^2\mathbf{b}(x). \tag{4.28}$$

with $p = \Phi'$. An equation for Φ' is obtained by taking the determinant of the matrix in (4.26). We define the Hamiltonian[1] as

$$\mathcal{H}(x,p) \equiv h_1(x,p)h_2(x,p) - \beta h_1(x,p) - x^2 h_2(x,p). \tag{4.29}$$

The equation for Φ' can then be expressed as $\mathcal{H}(x,\Phi'(x)) = 0$. In particular, for the discrete process we have

$$\begin{aligned}\mathcal{H}_{\rm disc}(x,p) = \frac{x^2}{\varphi^2}(e^{-\varphi p}-1)^2 + \frac{\sigma}{\varphi^2}(e^{\varphi p}-1)^2 \\ + \frac{x}{\varphi^2}(\sigma+1)(e^{-\varphi p}-1)(e^{\varphi p}-1) \\ - \frac{x}{\varphi}(\beta+x^2)(e^{-\varphi p}-1) \\ - \frac{1}{\varphi}(\beta\sigma+x^2)(e^{\varphi p}-1),\end{aligned} \tag{4.30}$$

and for the semi-continuous process

$$\begin{aligned}\mathcal{H}_{\rm sc}(x,p) = b_1(x)b_2(x)p^4 + (b_1(x)v_2(x) + b_2(x)v_1(x))p^3 \\ + (v_1(x)v_2(x) - (\beta b_1(x) + x^2(x)b_2(x)))p^2 \\ - (\beta v_1(x) + x^2 v_2(x))p.\end{aligned} \tag{4.31}$$

All of the roots of $\mathcal{H}_{\rm sc} = 0$ are real, but there is only one nontrivial root that vanishes at the deterministic fixed points, namely

$$\begin{aligned}\Phi'(x) = \frac{-1}{3b_1(x)b_2(x)}\Bigg[b_1(x)v_2(x) + b_2(x)v_1(x) + \frac{1}{2}\left(|Z(x)|^{1/3} + \frac{C_1(x)}{|Z(x)|^{1/3}}\right) \\ \times \left(\cos(\frac{1}{3}\arg(Z(x))) - \sqrt{3}\sin(\frac{1}{3}\arg(Z(x)))\right)\Bigg],\end{aligned} \tag{4.32}$$

where

$$Z(x) = \frac{1}{2}\left(C_2(x) + i\sqrt{4C_1(x)^3 - C_2(x)^2}\right), \tag{4.33}$$

$$\begin{aligned}C_1(x) = (b_1(x)v_2(x) + b_2(x)v_1(x))^2 \\ - 3b_1(x)b_2(x)[v_1(x)v_2(x) - (x^2 b_2(x) + \beta b_1(x))],\end{aligned} \tag{4.34}$$

[1] So named because of the similarities to classical Hamiltonian dynamics, even though it does not necessarily represent a physical energy. It does, however, define a scalar potential Φ.

and

$$\begin{aligned} C_2(x) = &-2(b_1(x)v_2(x) + b_2(x)v_1(x))^3 \\ &+ 9b_1(x)b_2(x)(b_1(x)v_2(x) + b_2(x)v_1(x)) \\ &\qquad \times [v_1(x)v_2(x) - (x^2 b_2(x) + \beta b_1(x))] \\ &+ 27(b_1(x)b_2(x))^2(x^2 v_2(x) + \beta v_1(x)). \end{aligned} \tag{4.35}$$

For the discrete problem, the Hamiltonian $\mathcal{H}_{\text{disc}}(x,p)$ can be transformed to a cubic polynomial in $q = e^{\varphi p}$. Then, the solutions are given by the positive real roots of the cubic polynomial

$$\begin{aligned} \frac{\varphi^2 q^2}{(q-1)}\mathcal{H}(x, \frac{\ln(q)}{\varphi}) = \sigma q^3 - (x + \varphi x^2 + \sigma(1 + x + \varphi\beta))q^2 \\ + x(1 + x + \varphi(\beta + x^2) + \sigma)q - x^2. \end{aligned} \tag{4.36}$$

There is a single suitable root, given by

$$\begin{aligned} q(x) = \frac{-1}{3\sigma}\Bigg[&- x(1+\varphi x) - \sigma(1 + x + \varphi\beta) - \frac{1}{2}\left(|Y(x)|^{1/3} + \frac{Q_1(x)}{|Y(x)|^{1/3}}\right) \\ &\times \left(\cos(\frac{1}{3}\arg(Y(x))) - \sqrt{3}\sin(\frac{1}{3}\arg(Y(x)))\right)\Bigg], \end{aligned} \tag{4.37}$$

where

$$Y(x) = \frac{1}{2}\left(Q_2(x) + i\sqrt{4Q_1(x)^3 - Q_2(x)^2}\right), \tag{4.38}$$

$$Q_1(x) = (x(1+\varphi x) + \sigma(1 + x + \varphi\beta))^2 - 3\sigma x(1 + x + \varphi(\beta + x^2) + \sigma), \tag{4.39}$$

and

$$\begin{aligned} Q_2(x) = 9\sigma x\left(\sigma\left(1 + x + \beta\varphi\right) + x\left(1 + \varphi x\right)\right)\left(1 + \sigma + x + \varphi\left(\beta + x^2\right)\right) \\ - 2\left(\sigma\left(1 + x + \beta\varphi\right) + x\left(1 + \varphi x\right)\right)^3 - 27\sigma^2 x^2. \end{aligned} \tag{4.40}$$

The potential function $\Phi(x)$ for both approximations is obtained numerically by quadrature. We rewrite the remaining term in (4.25) as

$$\mathbf{r}_0(x) = k(x)\hat{\mathbf{w}}(x), \tag{4.41}$$

where the approximation for the conditional internal state distribution (2.7) is determined by calculating the nullspace of the matrix $(A + \Sigma_{\mathbf{h}})$, which is

$$\hat{\mathbf{w}}(x) \equiv \begin{bmatrix} \frac{-h_2(x,\Phi'(x))}{h_1(x,\Phi'(x)) - h_2(x,\Phi'(x))} \\ \frac{h_1(x,\Phi'(x))}{h_1(x,\Phi'(x)) - h_2(x,\Phi'(x))} \end{bmatrix}. \tag{4.42}$$

The scalar function $k(x)$ is a normalization factor and is determined at higher order.

To calculate $k(x)$, substitute (4.25) into (4.20) and collect $O(\epsilon)$ terms to get

$$O(\epsilon): \quad \left[A + \Sigma_{\mathbf{h}(x,\Phi'(x))}\right] \mathbf{r}_1 = \frac{dk}{dx}\Sigma_{\mathbf{h}_p}\hat{\mathbf{w}} + k\left[\left(\Sigma_{\mathbf{h}_{px}} + \frac{1}{2}\Phi''(x)\Sigma_{\mathbf{h}_{pp}}\right)\hat{\mathbf{w}} + \Sigma_{\mathbf{h}_p}\frac{d\hat{\mathbf{w}}}{dx}\right].$$

The left nullvector, satisfying $\boldsymbol{l}^T[A + \Sigma_{\mathbf{h}(x,\Phi'(x))}] = 0$, is

$$\boldsymbol{l}(x) = \mathbf{1} - \begin{bmatrix} \frac{h_2(x,\Phi'(x))}{\beta+x^2} \\ \frac{h_1(x,\Phi'(x))}{\beta+x^2} \end{bmatrix}, \tag{4.43}$$

and it follows from the Fredholm Alternative theorem that $k(x)$ satisfies

$$\frac{dk}{dx} + \Psi'(x)k = 0, \tag{4.44}$$

where

$$\Psi'(x) = \frac{\boldsymbol{l}^T(x)\mathbf{H}_{px}(x,\Phi'(x)) + \frac{1}{2}\Phi''(x)\boldsymbol{l}^T(x)\mathbf{H}_{pp}(x,\Phi'(x))}{\boldsymbol{l}^T(x)\mathbf{H}_p(x,\Phi'(x))}, \tag{4.45}$$

for $x \neq x_\pm, x_*$, and

$$\mathbf{H}(x,p) \equiv [A + \Sigma_{\mathbf{h}(x,p)}]\hat{\mathbf{w}}(x). \tag{4.46}$$

We can express $\Phi''(x)$ in terms of partial derivatives of the Hamiltonian with

$$\Phi''(x) = -\frac{\mathcal{H}_x(x,\Phi'(x))}{\mathcal{H}_p(x,\Phi'(x))}, \quad x \neq x_\pm, x_*. \tag{4.47}$$

For more about evaluating the limit $x \to x_c$, $x_c = x_\pm, x_*$, of $\Phi''(x)$ and $\Psi'(x)$, see Appendix C. Integrating (4.44) gives

$$k(x) = \exp\left[-\Psi(x)\right], \tag{4.48}$$

where $\Psi(x)$ is computed by numerical quadrature of (4.45). Finally, to leading order, the eigenfunction is

$$\boldsymbol{\phi}_0(x) \sim \hat{\mathbf{w}}(x)k(x)\exp\left[-\frac{1}{\epsilon}\Phi(x)\right], \tag{4.49}$$

where $k(x)$ and $\hat{\mathbf{w}}(x)$ are defined by (4.48) and (4.42), respectively, while $\Phi(x)$ is determined by numerical quadrature[2] of $\Phi'(x)$, defined by (4.32) for the semi-continuous process and (4.37) for the discrete process. Note that at fixed points, $x_c = x_\pm, x_*$, we have that $\hat{\mathbf{w}}(x_c) = \boldsymbol{\rho}(x_c)$, where $\boldsymbol{\rho}(x)$ is defined by (4.19).

[2]In practice, we find that the best way of numerically integrating $\Phi'(x)$ and $\Psi'(x)$ is to use Chebychev approximation methods (we use the GNU Scientific Library).

4.1.1 Adiabatic limit

In the adiabatic limit, we have that $\hat{\mathbf{w}}(x) = \boldsymbol{\rho}(x)$ (see (4.19)). The quasi-stationary analysis is well known for the reduced process, and because of the reduction in the complexity of the process, a simple analytical approximation can be obtained. We quote the result here and refer the reader to [5, 32]; that is,

$$\Phi(x) = x\left(\ln(\frac{x}{f(x)}) - 1\right) + 2\sqrt{\beta}\left(\tan^{-1}(\frac{x}{\sqrt{\beta}}) - \sqrt{\sigma}\tan^{-1}(\frac{x}{\sqrt{\sigma\beta}})\right), \quad (4.50)$$

and

$$k(x) = \frac{1}{\sqrt{xf(x)}}, \quad (4.51)$$

where $f(x)$ is given by (3.9).

4.1.2 Quasi-deterministic limit

Both the discrete and semi-continuous process converge to the QD process (Section 3.2) in the limit $\varphi \to 0$ (i.e., $\alpha_e \to \infty$ with α_i fixed). A similar analysis can be carried out on the CK equation (3.5) for this process (see [16, 28, 30] for details). The result is a simple analytical approximation. For $\sigma < x < 1$ we have

$$\hat{\mathbf{w}}(x) = \begin{bmatrix} \frac{1-x}{1-\sigma} \\ \frac{x-\sigma}{1-\sigma} \end{bmatrix}, \quad (4.52)$$

$$\Phi(x) = -\beta\ln(1-x) - \sigma^2\ln(x-\sigma) - \frac{1}{2}(x-\sigma)^2 - 2\sigma(x-\sigma), \quad (4.53)$$

and

$$k(x) = \frac{1}{(x-\sigma)(1-x)}. \quad (4.54)$$

4.2 Singular perturbation approximation of the adjoint eigenfunction $\xi^{(a)}(x)$

The primary purpose of this paper is to develop the singular perturbation techniques to approximate the adjoint eigenfunction. The WKB method used in the previous section is a straightforward extension of those presented in [28] for the QD limit, and it provides only an approximation of the stationary density, not the timescale for metastable transitions (i.e., the principal eigenvalue λ). To get information about transition times we must have an approximation of the adjoint eigenfunction. We emphasize that singular perturbation methods are ideally suited to higher dimensional problems that display nonequilibrium stationary behavior.

To make the following analysis applicable to a more general class of Markov processes, we assume that there are M internal states; for the discrete and semi-continuous process $M = 2$. Here we assume that $\mathbf{h} \in \mathbb{R}^M$, and the transition-rate matrix A is $M \times M$. Because the boundary layer analysis of the semi-continuous process is very different from the discrete process, we treat the two cases separately.

4.2.1 Semi-continuous process

Up to terms exponentially small in ϵ, the first adjoint eigenfunction satisfies

$$\left[[A(x)]^T + \epsilon \Sigma_{\mathbf{v}(x)} \frac{d}{dx} + \epsilon^2 \Sigma_{\mathbf{b}(x)} \frac{d^2}{dx^2} \right] \boldsymbol{\xi}_0(x) = 0, \tag{4.55}$$

along with the absorbing boundary condition,

$$\boldsymbol{\xi}_0(0) = 0. \tag{4.56}$$

The outer solution, which does not satisfy the boundary condition, is exactly $\boldsymbol{\xi}_{\text{out}} \equiv \mathbf{1}$.

To obtain an approximate solution that also satisfies boundary conditions, we must rescale $x = x_* + \epsilon^\theta z$, for some $\theta > 0$. A reasonable first try is to take $\theta = 1$ so that $x = x_* + \epsilon z$. Equation (4.55) becomes

$$\left[[A(x_*)]^T + \Sigma_{\mathbf{v}(x_*)} \frac{d}{dz} + \Sigma_{\mathbf{b}(x_*)} \frac{d^2}{dz^2} \right] \boldsymbol{\xi}_{\text{bl}}(z) = 0. \tag{4.57}$$

The solution is a linear combination of the subsolutions $c_j \Upsilon_j e^{-\gamma_j z}$, $j = 0, \cdots, 2M-1$ where

$$\left(A(x_*)^T - \gamma_j \Sigma_{\mathbf{v}(x_*)} + \gamma_j^2 \Sigma_{\mathbf{b}(x_*)} \right) \Upsilon_j = 0, \tag{4.58}$$

and c_j, $j = 0, \cdots, 2M-1$ are unknown constants. We can specify the first solution as $\Upsilon_0 = \mathbf{1}$ and $\gamma_0 = 0$. We also require the solution to be bounded in the limit $z \to \infty$, which means that $c_j = 0$ if $\gamma_j < 0$; although we do not know apriori how many of the eigenvalues are negative. Note that the boundary condition (4.56) provides a system of M linear equations for the $2M$ unknowns, c_j, which means that constraints to eliminate the remaining M unknowns are required to close the system. One such constraint eliminates an unknown (i.e., c_0) by matching to the outer solution, leaving $M-1$ more constraints we must find, which suggests that there are $M-1$ negative eigenvalues. For simplicity, we order the eigenvalues so that $\gamma_j < 0$ for $j = M+1, \cdots, 2M-1$.

For the moment, consider the matrices in (4.58) as depending on x so that Υ_j and γ_j are also functions of x. It is simple to show that $\Upsilon_0(x) = \mathbf{1}$ and $\gamma_0(x) = 0$ even if the matrices are evaluated away from x_*. However, one of the solutions, label it $j = 1$, is $\gamma_1 = p(x) = \Phi'(x)$ from (4.32). Moreover, $\Upsilon_1(x) \to \mathbf{1}$ and $\gamma_1(x) \to 0$ as $x \to x_*$. In fact, $\gamma_1(x)$ vanishes at all of the deterministic fixed points since $\Phi'(x_c) = 0$, for $x = x_\pm, x_*$, and is nonzero at all other points.

It follows that the zero eigenvalue has a degenerate eigenspace, and the solution must include a secular term involving the generalized eigenvector satisfying

$$A(x_*)^T\boldsymbol{\zeta} = \Sigma_{\mathbf{v}(x_*)}\mathbf{1} = \mathbf{v}(x_*). \tag{4.59}$$

One can show [30] that the deterministic fixed points are the only points where this matrix has a degenerate eigenspace associated with the zero eigenvalue. The solution to (4.57) is thus

$$\boldsymbol{\xi}_{\mathrm{bl}}(z) = c_0\mathbf{1} + c_1(\boldsymbol{\zeta} - z\mathbf{1}) + \sum_{j=2}^{M} c_j\Upsilon_j e^{-\gamma_j z}. \tag{4.60}$$

However, because of the secular term, the solution remains unbounded in the limit $z \to \infty$, and as a result, it cannot be matched to the outer solution. Therefore, there is a transition layer which sits between the boundary layer and the outer region.

To find the scaling for this transition layer, we change variables to $x = \epsilon^\theta s$, for $0 < \theta < 1$, and define $\boldsymbol{\xi}_\theta(s) = \boldsymbol{\xi}_0(\frac{x}{\epsilon^\theta})$. Introduce the asymptotic expansion

$$\boldsymbol{\xi}_\theta(s) \sim \boldsymbol{\xi}_\theta^{(0)}(s) + \epsilon^\kappa\boldsymbol{\xi}_\theta^{(1)}(s) + \epsilon^{2\kappa}\boldsymbol{\xi}_\theta^{(2)}(s). \tag{4.61}$$

Equation (4.55) becomes

$$\begin{aligned}\Big[\big(A(x_*) + \epsilon^\theta sA'(x_*) + \cdots\big)^T \\ + \epsilon^{1-\theta}\left(\Sigma_{\mathbf{v}(x_*)} + \epsilon^\theta s\Sigma_{\mathbf{v}'(x_*)} + \cdots\right)\frac{d}{ds} \\ +\epsilon^{2(1-\theta)}\left(\Sigma_{\mathbf{b}(x_*)} + \cdots\right)\frac{d^2}{ds^2}\Big] \\ \times\left(\boldsymbol{\xi}_\theta^{(0)}(s) + \epsilon^\kappa\boldsymbol{\xi}_\theta^{(1)}(s) + \epsilon^{2\kappa}\boldsymbol{\xi}_\theta^{(2)}(s)\right) = 0.\end{aligned} \tag{4.62}$$

Setting $\epsilon = 0$ in (4.62) yields

$$O(1): \quad A(x_*)^T\boldsymbol{\xi}_\theta^{(0)}(s) = 0, \tag{4.63}$$

which implies that

$$\boldsymbol{\xi}_\theta^{(0)}(s) = a_0(s)\mathbf{1}, \tag{4.64}$$

for some scalar function $a_0(s)$. The leading-order terms in (4.62) then become

$$\begin{aligned}\epsilon^\kappa A(x_*)^T\boldsymbol{\xi}_\theta^{(1)}(s) + \epsilon^{1-\theta}a_0'(s)\mathbf{v}(x_*) \\ + O(\epsilon) + O(\epsilon^{\theta+\kappa}) + O(\epsilon^{2(1-\theta)}) + O(\epsilon^{2\kappa}) = 0.\end{aligned} \tag{4.65}$$

Note that $\frac{d^n}{dx^n} A^T \mathbf{1} = 0$ for all $n \geq 0$. Setting $\kappa = 1 - \theta$ we have

$$O(\epsilon^{1-\theta}): \quad A(x_*)^T \boldsymbol{\xi}_\theta^{(1)}(s) = -a_0'(s)\mathbf{v}(x_*), \tag{4.66}$$

and since $\boldsymbol{\rho}(x_*)^T \mathbf{v}(x_*) = 0$, the solution is

$$\boldsymbol{\xi}_\theta^{(1)}(s) = -a_0'(s)\boldsymbol{\zeta}, \tag{4.67}$$

where $\boldsymbol{\zeta}$ satisfies (4.59). Thus,

$$\boldsymbol{\xi}_\theta(s) \sim a_0(s)\mathbf{1} - \epsilon^{1-\theta} a_0'(s)\boldsymbol{\zeta}. \tag{4.68}$$

The function $a_0(s)$ is determined at higher order; we find

$$\begin{aligned} \epsilon^{1-\theta} A(x_*)^T \boldsymbol{\xi}_\theta^{(2)} + \epsilon^\theta a_0' \left(z\mathbf{v}'(x_*) - zA'(x_*)^T \boldsymbol{\zeta} \right) \\ - \epsilon^{1-\theta} a_0'' \left(\Sigma_{\mathbf{v}(x_*)} \boldsymbol{\zeta} - \mathbf{b}(x_*) \right) = 0. \end{aligned} \tag{4.69}$$

Setting $\theta = 1/2$ yields

$$O(\epsilon): \quad A(x_*)^T \boldsymbol{\xi}_\theta^{(2)}(s) = a_0''(s)\Sigma_{\mathbf{v}(x_*)}\boldsymbol{\zeta} - a_0'(s)s \left(\mathbf{v}'(x_*) - A'(x_*)^T \boldsymbol{\zeta} \right), \tag{4.70}$$

and the resulting solvability condition is

$$a_0''(s) - s \left(\frac{(\boldsymbol{\rho}(x_*)^T \mathbf{v}(x_*))'}{\boldsymbol{\rho}(x_*)^T \left(\Sigma_{\mathbf{v}(x_*)} \boldsymbol{\zeta} - \mathbf{b}(x_*) \right)} \right) a_0'(s) = 0. \tag{4.71}$$

Note that $A\boldsymbol{\rho} = 0 \Rightarrow A'\boldsymbol{\rho} = -A\boldsymbol{\rho}'$. Furthermore, $\boldsymbol{\zeta}^T A(x_*)\boldsymbol{\rho}'(x_*) = \mathbf{v}(x_*)^T \boldsymbol{\rho}'(x_*)$. Hence, $\boldsymbol{\rho}(x_*)^T \mathbf{v}'(x_*) - \boldsymbol{\rho}(x_*)^T A'(x_*)\boldsymbol{\zeta} = (\boldsymbol{\rho}(x_*)^T \mathbf{v}(x_*))'$. One can show that (see Appendix A)

$$\frac{\boldsymbol{\rho}(x_*)^T \mathbf{v}'(x_*)}{\boldsymbol{\rho}(x_*)^T (\mathbf{b}(x_*) - \Sigma_{\mathbf{v}(x_*)}\boldsymbol{\zeta})} = -\Phi''(x_*). \tag{4.72}$$

Assuming that $\Phi''(x_*) < 0$, the solution to (4.71) is

$$a_0'(s) = \hat{c}_1 e^{\frac{1}{2}\Phi''(x_*)s^2}, \tag{4.73}$$

and

$$a_0(s) = \hat{c}_0 + \hat{c}_1 \int_0^s e^{\frac{1}{2}\Phi''(x_*)y^2} dy, \tag{4.74}$$

where $\hat{c}_{0,1}$ are unknowns constants. Note that since $x = x_*$ is a local maxima of $\Phi(x)$, we assume that $\Phi''(x_*) < 0$ so that $a_0(s) \to 0$ as $s \to \infty$. The solution (4.68) becomes

$$\boldsymbol{\xi}_{1/2}(s) \sim \left(\hat{c}_0 + \hat{c}_1 \int_0^s e^{\frac{1}{2}\Phi''(x_*)y^2} dy \right) \mathbf{1} - \epsilon^{1/2} \hat{c}_1 e^{\frac{1}{2}\Phi''(x_*)s^2} \boldsymbol{\zeta}, \tag{4.75}$$

which replaces the first two terms in (4.60) (i.e., $\hat{c}_0\mathbf{1} + \hat{c}_1(\boldsymbol{\zeta} - z\mathbf{1})$). Notice that the solution is now bounded in the limit $z \to \infty$, which allows us to match it to the outer solution; we require $\lim_{s\to\infty} \boldsymbol{\xi}_{1/2}(z) = \mathbf{1}$ so that

$$\hat{c}_0 + \hat{c}_1 \int_0^\infty e^{\frac{1}{2}\Phi''(x_*)s^2} ds = \hat{c}_0 + \hat{c}_1 \sqrt{\frac{\pi}{2\left|\Phi''(x_*)\right|}} = 1, \tag{4.76}$$

and take

$$\hat{c}_0 = 1 - \hat{c}_1 \sqrt{\frac{\pi}{2\left|\Phi''(x_*)\right|}}. \tag{4.77}$$

As $s \to 0$,

$$\boldsymbol{\xi}_{1/2}(s) \sim (\hat{c}_0 + \hat{c}_1 s)\mathbf{1} - \epsilon^{1/2}\hat{c}_1\boldsymbol{\zeta}. \tag{4.78}$$

This matches with (4.60) if $c_0 = \hat{c}_0$ and $c_1 = -\epsilon^{1/2}\hat{c}_1$. A uniform solution, valid throughout the boundary layer and transition regions, is

$$\begin{aligned}\boldsymbol{\xi}_0(x) \sim \left[1 - \hat{c}_1\left(\sqrt{\frac{\pi}{2\left|\Phi''(x_*)\right|}} - \int_{x_*}^{(x-x_*)/\epsilon^{1/2}} e^{\frac{1}{2}\Phi''(x_*)y^2} dy\right)\right]\mathbf{1} \\ - \epsilon^{1/2}\hat{c}_1 e^{\frac{1}{2}\Phi''(x_*)(x-x_*)^2/\epsilon}\boldsymbol{\zeta} + \sum_{j=2}^{M} c_j \Upsilon_j e^{-\gamma_j(x-x_*)/\epsilon}.\end{aligned} \tag{4.79}$$

The remaining unknown constants c_j, $j = 1, \cdots, M$, are determined using the absorbing boundary condition (4.56), resulting in the linear system of equations,

$$\hat{c}_1\left(\sqrt{\frac{\pi}{2\left|\Phi''(x_*)\right|}}\mathbf{1} + \epsilon^{1/2}\boldsymbol{\zeta}\right) - \sum_{j=2}^{M} c_j \Upsilon_j = \mathbf{1}. \tag{4.80}$$

Using this general result, we can calculate the adjoint eigenfunction for the model problem. We have that $M = 2$ for the semi-continuous process (Section 2.1), and the solution to (4.80) is

$$\hat{c}_1 \sim \sqrt{\frac{2\left|\Phi''(x_*)\right|}{\pi}} + O(\epsilon^{1/2}), \tag{4.81}$$

$$c_2 \sim -\epsilon^{1/2}\sqrt{\frac{2\left|\Phi''(x_*)\right|}{\pi}}\left(\frac{\Delta^T\boldsymbol{\zeta}}{\Delta^T\Upsilon_2}\right) + O(\epsilon), \tag{4.82}$$

where $\Delta \equiv [1, -1]^T$. A simple calculation shows that

$$\boldsymbol{\zeta} = \frac{1 - x_*}{2\beta}\begin{bmatrix}\beta + 1 \\ \beta - 1\end{bmatrix}, \tag{4.83}$$

and

$$\Upsilon_2 = \frac{\beta + x_*^2}{\beta(x_* - \sigma)R} \begin{bmatrix} -\beta + \gamma_2(1 - x_*) + \frac{\varphi}{2}\gamma_2^2(1 + x_*) \\ -\beta \end{bmatrix}, \tag{4.84}$$

$$R = \gamma_2(1 - x_*) - \frac{\varphi}{2}\gamma_2^2(1 + x_*). \tag{4.85}$$

where Υ_2 is normalized so that

$$\gamma_2 \Upsilon_2^T \Sigma_{\mathbf{b}(x_*)} \boldsymbol{\rho}(x_*) = \Upsilon_2^T \Sigma_{\mathbf{v}(x_*)} \boldsymbol{\rho}(x_*) = 1. \tag{4.86}$$

4.2.2 Discrete process

The adjoint eigenfunction for the discrete process satisfies

$$\left([A(x)]^T + \Sigma_{\mathbb{d}^*}\right) \boldsymbol{\xi}_0(x) = 0, \tag{4.87}$$

where A is given by (4.17); the adjoint operator is

$$\mathbb{d}^* = \frac{1}{\varphi} \begin{bmatrix} x(\mathbb{e}^{-\partial x} - 1) + \sigma(\mathbb{e}^{\partial x} - 1) \\ x(\mathbb{e}^{-\partial x} - 1) + (\mathbb{e}^{\partial x} - 1) \end{bmatrix}, \tag{4.88}$$

where $\mathbb{e}^{\pm\partial x}$ are defined by (4.22). The absorbing boundary condition is

$$\boldsymbol{\xi}_0(n_*) = 0, \tag{4.89}$$

Once again, the outer solution is $\boldsymbol{\xi}_{\text{out}} = \mathbf{1}$.

Motivated by the boundary layer analysis of the previous section, we rescale with $x = x_* + \epsilon^\theta s$. We are interested in two cases: $\theta = 1$ and $\theta = 1/2$. In the former case, the scaling simply returns the process to a discrete variable since $x = \varphi\epsilon n$. Let $\hat{n} = n - n_*$ and $\boldsymbol{\xi}_b(\hat{n}) = \boldsymbol{\xi}_0(x_* + \varphi\epsilon\hat{n})$. Then to leading order

$$\varphi[A(x_*)]^T \boldsymbol{\xi}_b(\hat{n}) + x_* \left(\boldsymbol{\xi}_b(\hat{n} - 1) - \boldsymbol{\xi}_b(\hat{n})\right) + \Sigma_{\mathbf{v}(0)} \left(\boldsymbol{\xi}_b(\hat{n} + 1) - \boldsymbol{\xi}_b(\hat{n})\right) = 0. \tag{4.90}$$

Solutions have the form

$$\boldsymbol{\xi}_b(\hat{n}) = \Upsilon_j \mu_j^{\hat{n}}. \tag{4.91}$$

Substituting this into (4.90) yields

$$\left[\varphi\mu_j[A(x_*)]^T + \mu_j(\mu_j - 1)\Sigma_{\mathbf{v}(0)} + x_*(\mu_j - 1)I\right] \Upsilon_j = 0, \tag{4.92}$$

with μ_j determined by the characteristic equation,

$$\det\left[\varphi\mu_j[A(x_*)]^T + \mu_j(\mu_j - 1)\Sigma_{\mathbf{v}(0)} + x_*(\mu_j - 1)I\right] = 0. \tag{4.93}$$

Just as in the previous section, one of the linearly independent solutions is

$$\boldsymbol{\xi}_b(\hat{n}) = \boldsymbol{\zeta} - \varphi\hat{n}\mathbf{1}, \tag{4.94}$$

where $\boldsymbol{\zeta}$ is given by (4.59).

On the other hand, if $\theta = 1/2$ we recover (4.62), which means that we can use (4.75) as a solution instead of (4.94). The uniformly valid solution is then

$$\boldsymbol{\xi}_0(x) \sim \left[1 - \hat{c}_1 \left(\sqrt{\frac{\pi}{2\left|\Phi''(x_*)\right|}} - \int_0^{\frac{x-x_*}{\epsilon^{1/2}}} e^{\frac{1}{2}\Phi''(x_*)y^2} dy\right)\right] \mathbf{1} - \epsilon^{1/2}\hat{c}_1 e^{\frac{1}{2}\Phi''(x_*)\frac{(x-x_*)^2}{\epsilon}} \boldsymbol{\zeta} + \sum_{j=2}^{M} c_j \Upsilon_j \mu_j^{\frac{x-x_*}{\varphi\epsilon}}, \quad (4.95)$$

where we assume that $|\mu_j| < 1$, $j = 2, \cdots, M$. The unknown constants are determined by the boundary condition $\boldsymbol{\xi}_0(x_*) = 0$, which results in the same linear system (4.80) as obtained for the semi-continuous process, except that Υ_j now satisfies (4.92).

Turning to the model problem, calculation of $\boldsymbol{\xi}_0(x_*)$ for the discrete process is very similar to the semi-continuous process. Using (4.81), (4.82) and (4.83) in (4.95), we need only calculate Υ_2 and μ_2 satisfying (4.93) and (4.93). We find that

$$\Upsilon_2 = \frac{\beta + x_*^2}{x_*^2 R}\begin{bmatrix} -\varphi x_*^2 \mu_2 \\ (\sigma\mu_2^2 - (\sigma + x_*)\mu_2 + x_*) - \varphi x_*^2 \mu_2 \end{bmatrix}, \quad (4.96)$$

$$R = (\sigma\mu_2^2 - (\sigma + x_*)\mu_2 + x_*) - \varphi(\beta\sigma + x_*^2)\mu_2, \quad (4.97)$$

and

$$\mu_2 = \frac{-\varphi(1-\sigma)x_*^3 - (\sigma+1)(x_* - \sigma)x_* + x_*\sqrt{U}}{2\sigma(\sigma - x_*)}, \quad (4.98)$$

$$U = (\sigma-1)(\varphi^2(\sigma-1)x_*^4 - 2\varphi(\sigma+1)x_*^3) + (\sigma-1)((2\sigma^2\varphi + (1+2\varphi)\sigma - 1)x_*^2 - 2\sigma(\sigma-1)x_* + \sigma^2(\sigma-1)). \quad (4.99)$$

4.3 Principal eigenvalue

Now that we have approximations for the right and left eigenfunction, we can construct the approximation of the principal eigenvalue using the spectral projection method (see (4.15)) outlined in the introduction of this section. For brevity, we only show steps for computing the eigenvalue corresponding to the right well for which the minimum is x_+. The corresponding result for the left well proceeds similarly, and the final formula differs only by substitution of x_- for x_+.

First, for the semi-continuous process, substituting the eigenfunctions (4.49)

and (4.79) into (4.14) yields

$$J(\boldsymbol{\phi}_0, \boldsymbol{\xi}_0) \sim \sqrt{\frac{2\epsilon\,|\Phi''(x_*)|}{\pi}} \left[\mathbf{b}(x_*)^T \boldsymbol{\rho}(x_*) + \left(\frac{\Delta^T \boldsymbol{\zeta}}{\Delta^T \Upsilon_2} \right) \gamma_2 \Upsilon_2^T \Sigma_{\mathbf{b}(x_*)} \boldsymbol{\rho}(x_*) \right] \times k(x_*) \exp\left[-\frac{1}{\epsilon} \Phi(x_*) \right] \tag{4.100}$$

$$= \sqrt{\frac{2\epsilon\,|\Phi''(x_*)|}{\pi}} \left[\langle \mathbf{b}(x_*) \rangle + D(x_*) \right] k(x_*) \exp\left[-\frac{1}{\epsilon} \Phi(x_*) \right],$$

where $\boldsymbol{\rho}$, $\boldsymbol{\zeta}$, and Υ_2 are given by (4.19), (4.83), and (4.84), respectively. Notice that

$$\frac{\Delta^T \boldsymbol{\zeta}}{\Delta^T \Upsilon_2} \gamma_2 \Upsilon_2^T \Sigma_{\mathbf{b}(x_*)} \boldsymbol{\rho}(x_*) = \frac{(x_* - \sigma)(1 - x_*)}{\beta + x_*^2} = D(x_*), \tag{4.101}$$

the diffusivity from the QSS reduction (2.21) evaluated at the unstable fixed point. (Clearly, the boundary term $J(\boldsymbol{\phi}_0, \boldsymbol{\xi}_0)$ is related to the classical diffusive flux $J(\hat{u}) = D \frac{d\hat{u}}{dx}$.) The normalization constant is well approximated by

$$\langle \boldsymbol{\phi}_0, \boldsymbol{\xi}_0 \rangle \sim \left(\frac{\Phi''(x_0)}{2\pi\epsilon} \right)^{-1/2}. \tag{4.102}$$

Finally, the eigenvalue approximation is

$$\lambda_0 \sim \left(\frac{B}{\pi} \sqrt{|\Phi''(x_*)|\,\Phi''(x_\pm)} \right) \frac{k(x_*)}{k(x_\pm)} \exp\left[-\frac{1}{\epsilon} (\Phi(x_*) - \Phi(x_\pm)) \right], \tag{4.103}$$

where

$$B = \varphi x_* + \frac{(x_* - \sigma)(1 - x_*)}{\beta + x_*^2}. \tag{4.104}$$

The preexponential factor is equivalently

$$\frac{k(x_*)}{k(x_\pm)} = \exp\left[-(\Psi(x_*) - \Psi(x_\pm)) \right]. \tag{4.105}$$

Recall that $\Phi(x)$ and $\Psi(x)$ are determined by numerical integration of (4.32) and (4.45), respectively.

The discrete version of (4.15) can be obtained using a summation by parts argument. The resulting boundary contribution is

$$J(\boldsymbol{\phi}_0, \boldsymbol{\xi}_0) = \boldsymbol{\phi}_0(x_*)^T \Sigma_{\mathbf{v}(0)} \boldsymbol{\xi}_0(x_* + 1/\alpha_\mathrm{e}). \tag{4.106}$$

The first two terms in $\boldsymbol{\xi}_0(x_* + 1/\alpha_\mathrm{e})$ (see (4.95)) can be expanded in $1/\alpha_\mathrm{e} \ll 1$

(the third term is the boundary layer solution), and the result is

$$J(\boldsymbol{\phi}_0, \boldsymbol{\xi}_0) \sim \boldsymbol{\phi}_0(x_*)^T \Sigma_{\mathbf{v}(0)} \left(\epsilon^{1/2} \hat{c}_1 (-\boldsymbol{\zeta} + \varphi \mathbf{1}) + c_2 \mu_2 \Upsilon_2 \right) \tag{4.107}$$
$$= \epsilon^{1/2} \hat{c}_1 \left[\varphi x_* + (\mu_2 - 1) \left(\frac{\Delta^T \boldsymbol{\zeta}}{\Delta^T \Upsilon_2} \right) \Upsilon_2^T \Sigma_{\mathbf{v}(0)} \boldsymbol{\rho}(x_*) \right]$$
$$\times k(x_*) \exp \left[-\frac{1}{\epsilon} \Phi(x_*) \right],$$

where Υ_2 is given by (4.96). Similar to (4.101), we find that

$$(\mu_2 - 1) \left(\frac{\Delta^T \boldsymbol{\zeta}}{\Delta^T \Upsilon_2} \right) \Upsilon_2^T \Sigma_{\mathbf{v}(0)} \boldsymbol{\rho}(x_*) = D(x_*). \tag{4.108}$$

The eigenvalue for the discrete process has the form (4.103) with the following substitutions. The potential function $\Phi(x)$ is determined by numerical integration of $\Phi'(x) = \frac{\ln(q(x))}{\varphi}$, where $q(x)$ is given by (4.37), and $\Psi(x)$ is determined by integration of (4.45), with $\mathbf{h}(x, p)$ given by (4.27).

5 Results

In this section, we compare the approximations of the stability landscape, defined as $-\epsilon \ln(\hat{u}(x))$, and of the mean time of a metastable transition from the minimum of one well to another. The shape of the stability landscape can be described as a double-well potential, and in Fig. 2 it is shown for $\sigma = 0.015$ and two different values of the bifurcation parameter, β, located within the region of deterministic bistability (see Section 3.1). The stability landscape is shown in two columns of plots, each using different parameter values. In the left column $\beta = 0.24$, which is near the bifurcation point that eliminates the right stability well, and in the right column $\beta = 0.11$, which is near the bifurcation eliminating the left stability well. Each row shows a different value of ϵ with $\varphi = 1$ so that both noise sources are present. Approximations of the stability landscape are given by $\Phi(x) + \epsilon\Psi(x)$, where Φ and Ψ are defined in Section 4.1. Note that the WKB approximation of the discrete process brakes down as $x \to 0$ due to small copy number, requiring a boundary correction (see Appendix E). Each approximation—the QSA discrete and semi-continuous approximations, and the diffusion approximation—is compared to a numerical approximation obtained by SVD decomposition in the top two rows for which $0 < \epsilon \ll 1$. In the bottom row we take the limit $\epsilon \to 0$. Note that the SVD approximation cannot be computed for this case. First, we observe that the QSA approximation of the discrete and semi-continuous process are so close that they are indistinguishable for every parameter set. (Indeed, we find this to be the case for all of the results presented in this paper). On the other hand, the diffusion approximation shows significant inaccuracies, particularly in the left stability well. The most significant aspect of the stability landscape that affects metastable transitions

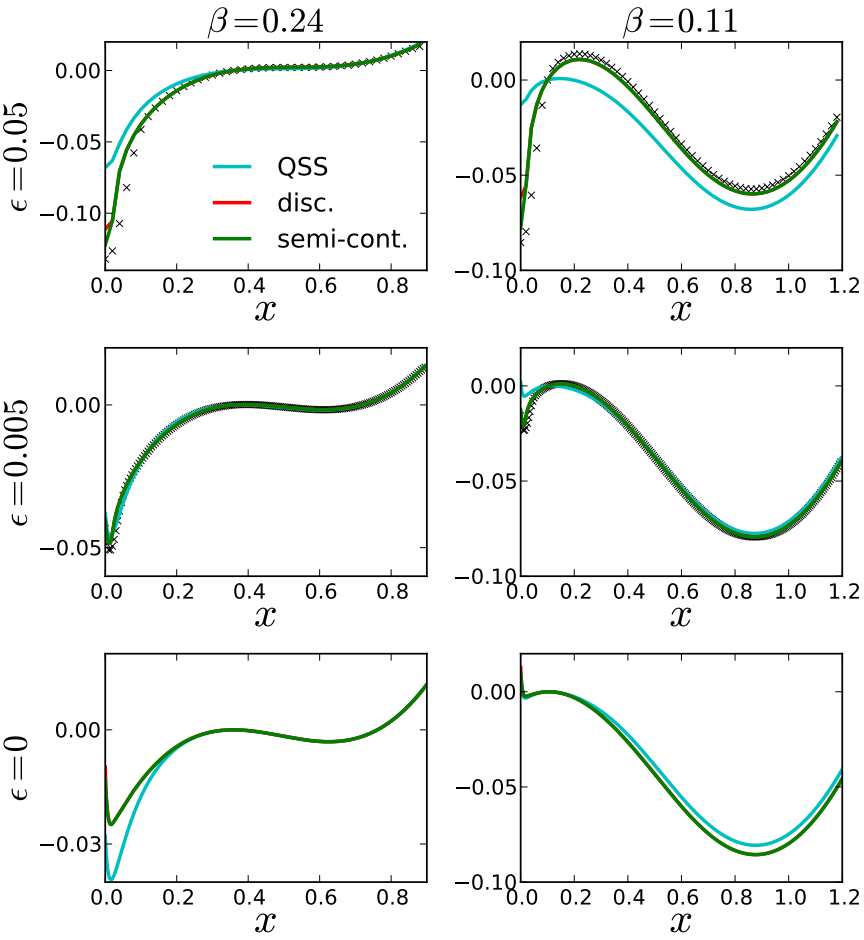

Figure 2: The stability landscape, $-\epsilon \log(\hat{u}(x))$, for $\varphi = 1$ and $\sigma = 0.015$. Each row shows results for a different value of ϵ, and each column shows a different value of the bifurcation parameter β. The light blue curve is the quasi-steady state approximation, the green curve is the semi-continuous QSA approximation, and the red curve (which cannot be seen beneath the green curve because both approximations are very close) is the discrete QSA approximation. For nonzero ϵ (the top four panes), the first and second order approximation $\Phi(x) + \epsilon\Psi(x)$ is compared to the value of $-\epsilon \log(p_{\rm s})$ obtained by a numerical SVD decomposition, shown as 'x' symbols. For $\epsilon = 0$ (the bottom two panes), the leading order approximation $\Phi(x)$ is shown. Note that the SVD solution can not be computed in the $\epsilon \to 0$ limit.

is the height of each well in the $\epsilon \to 0$ limit. Although the diffusion approximation does show some error in right stability well, including the height when $\epsilon = 0$, these differences are much less significant than the differences in the left well region. Even for the left well, the diffusion approximation is not always inaccurate. Indeed, all of the approximations closely agree when $\epsilon = 0.005$ and $\beta = 0.24$ (first column, second row of Fig. 2). However, for other values of ϵ (top and bottom row) this is clearly not the case.

To examine the differences in the approximations more closely, we plot the absolute error in the stability landscape and the error in the conditional internal state distribution in Fig. 3 for the parameter values used in the left column of

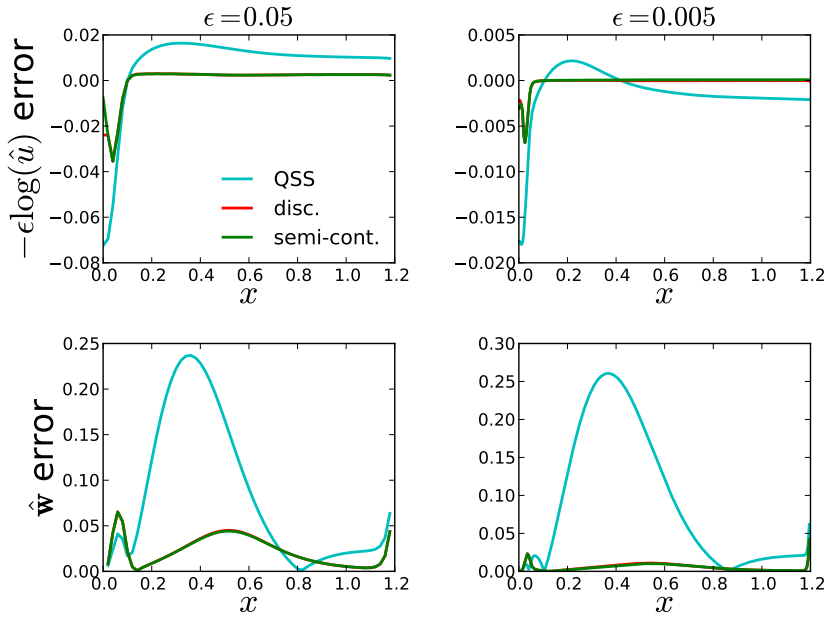

Figure 3: Absolute error. A comparison of each approximation to the numerical SVD result, for $\beta = 0.11$, $\sigma = 0.015$ shown in Fig. 2 (right column). The top row shows the error in stability landscape $-\epsilon \log(\hat{u}(x))$ for $\epsilon = 0.01$ and $\epsilon = 0.005$. The bottom row shows the error in the internal state distribution for the same values of ϵ. The colors for each curve are the same as in Fig. 2.

Fig. 2 (i.e., $\sigma = 0.015$, $\varphi = 1$, and $\beta = 0.11$). The conditional internal state distribution $\hat{\mathbf{w}}(x)$ is (4.42) for the discrete and semi-continuous QSA approximations and (4.19) for the QSS approximation. These are again compared to a numerical approximation obtained using an SVD decomposition, and the error is measured using the 1-norm (i.e., $\sum_{s=0}^{1} |[\hat{\mathbf{w}}_{\text{svd}}(x)]_s - [\hat{\mathbf{w}}_{\text{approx}}(x)]_s|$). The discrete and semi-continuous QSA approximations of the stability landscape show errors primarily in the left well region, while the QSS approximation also shows

some error in the right well. Interestingly, the conditional internal state distribution error is significant for the QSS approximation, peaking at 25% between x_* and x_+. We expect this error to be quite small near the deterministic fixed points, where all the approximations agree. We emphasize as one of the key results of this paper that away from fixed points, the conditional internal state distribution is *not* always close to the steady-state distribution as assumed in the QSS approximation method. This has been shown rigorously for velocity jump processes [30], for which the QD limit is an example.

The approximation of the mean time for a metastable transition between wells is shown in Fig. 4. The mean escape time approximations, defined as

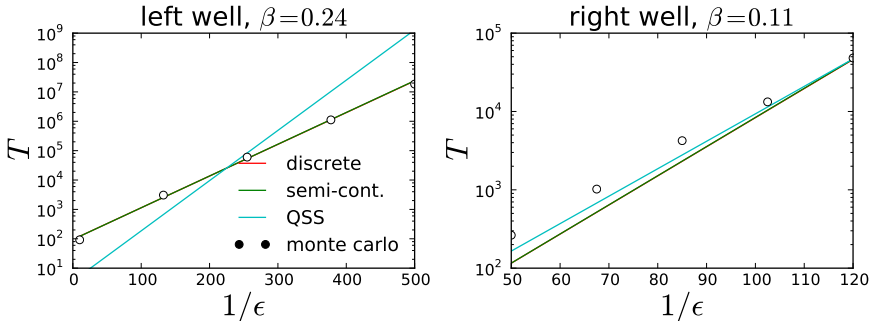

Figure 4: Mean exit time approximations compared to Monte-Carlo results. (a) Exit from the left well for the same parameters as used in the first column of Fig. 2. (b) Exit from the right well for the same parameters as used in the second column of Fig. 2.

$T \sim 1/\lambda$ (see (4.103)), are compared to exact Monte-Carlo (MC) simulations for parameter values used in Fig. 2. The mean escape time is plotted as a function of $1/\epsilon$ because each approximation is a linear function of this quantity, with a slope determined by the height of the potential well in the $\epsilon \to 0$ limit (see Fig. 2 bottom row). Escape from the left well (for $\beta = 0.11$, Fig. 2 left column) is shown on the left, where the discrete and semi-continuous QSA approximations are in good agreement with MC simulations. The three approximations converge near $\epsilon = 0.005$ consistent with Fig. 2 (first column, second row).

A somewhat unexpected result is obtained for escape from the right well (corresponding to the right column of Fig. 2). All three approximations are very close, and the QSS approximation is actually more accurate for smaller values of $1/\epsilon$. The difference in the slope of each approximation is slight (see Fig. 2 right column, bottom row) and the error in the QSS approximation should grow as $1/\epsilon \to \infty$. We cannot offer a definitive explanation for the accuracy of the QSS approximation for escape from the right well. One explanation consistent with our results is that the QSS approximation is valid for larger values of x, which seems reasonable since it relies on fast transitions between internal states and the rate of transitioning from the inactive to the active internal state is proportional to x^2. However, this is inconsistent with the error in the conditional internal

state distribution shown in Fig. 3 (bottom row), which is the key assumption underlying the QSS approximation.

Finally, we compare the mean escape time in the adiabatic limit $\alpha_\mathrm{i} \to \infty$ and in the QD limit $\alpha_\mathrm{e} \to \infty$. In Fig. 5, the mean time for escape from the left

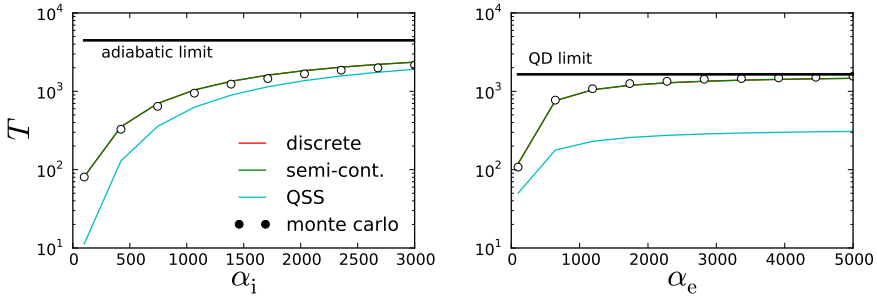

Figure 5: The mean exit time for escape from the left well to the right well, with $\beta = 0.23$, $\sigma = 0.04$. Three different approximations (solid curves) are compared to Monte-Carlo simulation results (symbols). The discrete (red) and semi-continuous (green) QSA approximations are indistinguishable. Also shown is the QSS approximation (light blue). (a) The mean exit time as a function of α_e for fixed $\alpha_\mathrm{i} = 333$. (b) The mean exit time as a function of α_i for fixed $\alpha_\mathrm{e} = 200$.

well is shown for $\sigma = 0.04$ and $\beta = 0.23$. In contrast to previous results, we do not fix $\varphi = \alpha_\mathrm{i}/\alpha_\mathrm{e} = 1$. Fig. 5 (right) illustrates that the discrete and semi-continuous approximation converge in the QD limit, and as expected, the QSS approximation error is significant. In the adiabatic limit (Fig. 5 left) the discrete and semi-continuous QSA approximations show close agreement for all values of α_i, and as expected, all three approximations converge as $\alpha_\mathrm{i} \to \infty$. Even though the discrete and semi-continuous QSA approximations do not converge in the adiabatic limit, the difference is very small. This suggests that a diffusion approximation for the external state—recall that we used such a procedure to derive the semi-continuous process from the full discrete process—may be valid in certain situations, which is interesting since diffusion approximations generally break down for metastable behavior due to large deviation errors. It is possible that the good agreement that we see for the example problem is due to the linear nature of the birth-death process governing transitions in the external state (i.e., that it is due to the simplicity of the example problem). Since the general QSA procedure presented here does not depend on this assumption, it would be interesting to see how this type of diffusion approximation behaves if the external state is governed by a more complicated process.

A Curvature prefactor

The purpose of this section is to show that the part of the eigenvalue estimate that contains information about the curvature of the stability well at the stable and unstable fixed point is unaffected by the diffusion approximations in Sections 2.1 and 2.2. This is a reflection of the fact that diffusion approximations, in general, are accurate in a neighborhood of deterministic fixed points. Each eigenvalue approximation (4.103) and (2.26) contains a prefactor term of the form $\sqrt{|\Phi''(x_*)|\,\Phi''(x_\pm)}$. We would like to show that, when evaluated at a fixed point, call it x_c, the second derivative of the potential function for the discrete, semi-continuous, and QSS processes are all identical. Clearly, evaluating the derivative in the first two cases is messy. However, we can also express the second derivative in terms of the Hamiltonian functions (4.53) and (4.30) as follows.

Differentiating $\mathcal{H}(x, \Phi'(x)) = 0$ with respect to x yields

$$\frac{d}{dx}\mathcal{H}(x, \Phi'(x)) = \mathcal{H}_x(x, \Phi'(x)) + \Phi''(x)\mathcal{H}_p(x, \Phi'(x)) = 0, \tag{A.1}$$

and it follows that

$$\Phi''(x) = -\frac{\mathcal{H}_x(x, \Phi'(x))}{\mathcal{H}_p(x, \Phi'(x))}. \tag{A.2}$$

However, we have that

$$\mathcal{H}_p(x_c, 0) = \mathcal{H}_x(x_c, 0) = \mathcal{H}_{xx}(x_c, 0) = 0. \tag{A.3}$$

A formula valid at fixed points can be obtained as follows. Differentiating $\mathcal{H}(x, \Phi'(x)) = 0$ twice with respect to x yields

$$\frac{d^2}{dx^2}\mathcal{H}(x, \Phi'(x)) = \mathcal{H}_{xx} + \Phi''\mathcal{H}_{xp} + \Phi''(\mathcal{H}_{px} + \Phi''\mathcal{H}_{pp}) + \Phi'''\mathcal{H}_p = 0, \tag{A.4}$$

and it follows from (A.3) that

$$\Phi''(x_c) = \frac{-2\frac{\partial^2}{\partial p \partial x}\mathcal{H}(x_c, 0)}{\frac{\partial^2}{\partial p^2}\mathcal{H}(x_c, 0)}. \tag{A.5}$$

A straightforward calculation shows that Taylor expansions of both Hamiltonian functions (4.31) and (4.30) agree to $O(p^2)$, and that

$$\Phi''(x_c) = -\frac{\frac{d}{dx}\langle \mathbf{v}(x_c)\rangle}{\frac{\langle \mathbf{b}(x_c)\rangle}{2\alpha_e \epsilon} + D(x_c)}, \tag{A.6}$$

which is also the second derivative of (2.25) since $\langle \mathbf{v}(x_c)\rangle = 0$.

B Adiabatic limit of the discrete process

Consider the Master equation for the probability distribution function

$$\mathrm{p}_j(\mathbf{n},t) \equiv \mathrm{p}(j,\mathbf{n},t|j_0,\mathbf{n}_0,t_0).$$

In matrix/operator form, the CK equation is

$$\frac{d\mathbf{p}}{dt} = L_1 p + \frac{1}{\epsilon} L_2 p, \tag{B.1}$$

where $\mathbf{p}(\mathbf{n},t) = (\mathrm{p}_1(\mathbf{n},t),\ \mathrm{p}_2(\mathbf{n},t), \cdots,\ \mathrm{p}_M(\mathbf{n},t))^T$; $L_1 = \Sigma_{\mathcal{D}_j}$ is a diagonal matrix of linear operators acting on $\mathbf{n}$, each of which has a $\mathbb{W}$-matrix representation; L_2 is an $M \times M$ $\mathbb{W}$-matrix governing the transitions between internal states, with transition rates that may depend on $\mathbf{n}$. Define the projection operator

$$\mathcal{P} \equiv \boldsymbol{\rho}\mathbf{1}^T, \tag{B.2}$$

where $L_2\boldsymbol{\rho}(\mathbf{n}) = 0$, with $\boldsymbol{\rho}(\mathbf{n}) > 0$ and $\sum_{j=1}^{M} \rho_j(\mathbf{n}) = 1$; and $\mathbf{1} \equiv (1,\ 1, \cdots,\ 1)^T$. We assume the solution has the following form

$$\mathbf{p}(\mathbf{n},t) = \mathcal{P}\mathbf{p}(\mathbf{n},t) + (I-\mathcal{P})\mathbf{p}(\mathbf{n},t) = u(\mathbf{n},t)\boldsymbol{\rho}(\mathbf{n}) + \epsilon\mathbf{w}(\mathbf{n},t), \tag{B.3}$$

where

$$u(\mathbf{n},t) \equiv \mathbf{1}^T\mathbf{p}(\mathbf{n},t), \quad \mathbf{1}^T\mathbf{w}(\mathbf{n},t) = 0. \tag{B.4}$$

Applying the projection operator to both sides of (B.1) yields

$$\frac{du}{dt}\boldsymbol{\rho} = \mathcal{P}L_1(u\boldsymbol{\rho} + \epsilon\mathbf{w}). \tag{B.5}$$

On the other hand, applying the orthogonal projection yields

$$\epsilon\frac{d\mathbf{w}}{dt} - \epsilon(I-\mathcal{P})L_1\mathbf{w} = (I-\mathcal{P})L_1(u\boldsymbol{\rho}) + L_2\mathbf{w}. \tag{B.6}$$

After setting $\epsilon = 0$ in the above equation we get

$$\mathbf{w}(\mathbf{n},t) \sim -L_2^{-1}(I-\mathcal{P})L_1(u(\mathbf{n},t)\boldsymbol{\rho}(\mathbf{n})). \tag{B.7}$$

Substituting (B.7) into (B.5) yields the scalar-valued operator equation for $u(\mathbf{n},t)$

$$\frac{du}{dt} = \mathbf{1}^T L_1(u\boldsymbol{\rho}) - \epsilon\mathbf{1}^T L_1 L_2^{-1}(I-\mathcal{P})L_1(u\boldsymbol{\rho}). \tag{B.8}$$

One can rewrite (B.8) in matrix form to obtain a linear system of ODEs for the vector $\mathbf{u}(t)$ with elements $u_{\mathbf{n}}(t) \equiv u(\mathbf{n},t)$

$$\frac{d\mathbf{u}}{dt} = W\mathbf{u}, \tag{B.9}$$

where

$$W \equiv \sum_{j=1}^{M} \mathcal{D}_j \rho_j. \tag{B.10}$$

Note that the operators $\mathcal{D}_j$ are now explicitly assumed to be matrices. The resulting transition rate matrix, W, reflects the reduced graph structure. The transition rates contained in W can be interpreted as the average of the transition rates contained in $\mathcal{D}_j$ with respect to the steady state distribution ρ_j. Note that since $\boldsymbol{\rho} > 0$ and $\mathcal{D}_j$ are $\mathbb{W}$-matrices, it follows that W is a $\mathbb{W}$-matrix. In general, the reduced equation represents a Markov process only at leading order.

C WKB/KM expansion

Consider the action of the operator $\mathfrak{e}^{\partial x}$ on $g(x)e^{-\alpha_{\mathrm{e}}\tilde{\Phi}(x)}$ where $g(x)$ is scalar function and $\tilde{\Phi}(x) = \varphi\Phi(x)$. We have that

$$\begin{aligned}
&\mathfrak{e}^{\pm\partial x}\left(g(x)e^{-\alpha_{\mathrm{e}}\tilde{\Phi}(x)}\right) \\
&= \sum_{n=0}^{\infty} \frac{(\pm 1)^n}{n!\alpha_{\mathrm{e}}^n} \frac{d^n}{dx^n}\left[g(x)e^{-\alpha_{\mathrm{e}}\tilde{\Phi}(x)}\right] \\
&= \sum_{n=0}^{\infty} \frac{(\pm 1)^n}{n!\alpha_{\mathrm{e}}^n} \sum_{k=0}^{n} \binom{n}{k} g^{(n-k)}(x) \frac{d^k}{dx^k} e^{-\alpha_{\mathrm{e}}\tilde{\Phi}(x)} \\
&= \sum_{n=0}^{\infty} \frac{(\pm 1)^n}{n!\alpha_{\mathrm{e}}^n} \sum_{k=0}^{n} \binom{n}{k} g^{(n-k)}(x)(1 + \mathrm{B}_k(-\alpha_{\mathrm{e}}\tilde{\Phi}(x)))e^{-\alpha_{\mathrm{e}}\tilde{\Phi}(x)},
\end{aligned} \tag{C.1}$$

where B_k is the kth complete Bell polynomial,

$$\mathrm{B}_n(f(x)) \equiv \det \begin{bmatrix} f' & \binom{n-1}{1} f'' & \binom{n-1}{2} f^{(3)} & \cdots & f^{(n)} \\ -1 & f' & \binom{n-2}{1} f'' & \cdots & f^{(n-1)} \\ 0 & -1 & f' & \cdots & f^{(n-2)} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & f' \end{bmatrix}, \tag{C.2}$$

and $B_0 = 0$. One can show that

$$\mathrm{B}_k(-\alpha_{\mathrm{e}}\tilde{\Phi}(x)) = \alpha_{\mathrm{e}}^k(-\tilde{\Phi}'(x))^k - \alpha_{\mathrm{e}}^{k-1}\frac{k}{2}(k-1)\tilde{\Phi}''(x)(-\tilde{\Phi}'(x))^{k-2} + O(\alpha_{\mathrm{e}}^{k-2}). \tag{C.3}$$

Expanding (C.1) in terms of $1/\alpha_{\rm e}$ yields

$$\begin{aligned}&\mathfrak{e}^{\pm\partial x}\left(g(x)e^{-\alpha_{\rm e}\tilde{\Phi}(x)}\right)\\&\quad=e^{\mp\tilde{\Phi}'(x)}\left[g(x)-\frac{1}{\alpha_{\rm e}}\left(g'(x)\mp\frac{1}{2}g(x)\tilde{\Phi}''(x)\right)+O(\alpha_{\rm e}^{-2})\right]e^{-\alpha_{\rm e}\tilde{\Phi}(x)}.\end{aligned}\tag{C.4}$$

D Evaluating $\Psi'(x_c)$ for $x_c = x_\pm, x_*$

Using L'Hôpital's rule, we find that

$$\begin{aligned}\Psi'(x_c)=\Bigg[&H_{pxx}+\frac{1}{2}\Phi''(x_c)(3H_{ppx}+\Phi''(x_c)H_{ppp})+\frac{1}{2}\Phi'''(x_c)H_{pp}\\&+\boldsymbol{l}'(x_c)^T\mathbf{H}_{px}(x_c,0)+\frac{1}{2}\Phi''(x_c)\boldsymbol{l}'(x_c)^T\mathbf{H}_{pp}(x_c,0)\Bigg]\\&\Big/\Big[\boldsymbol{l}'(x_c)^T\mathbf{H}_{p}(x_c,0)+H_{px}+\Phi''(x_c)H_{pp}\Big],\end{aligned}\tag{D.1}$$

where $\mathbf{H}(x,p)$ is defined by (4.46) and partial derivatives of $H(x,p)\equiv\mathbf{1}^T\mathbf{H}(x,p)$ are evaluated at $x=x_c$ and $p=0$, as for example,

$$H_{xp}\equiv\mathbf{1}^T\frac{\partial^2}{\partial x\partial p}\mathbf{H}(x_c,0).\tag{D.2}$$

We also have that $\Phi''(x_c)$ is given by (A.5), and

$$\Phi'''(x_c)=-2\frac{\mathcal{H}_{pxx}(x_c,0)+\frac{1}{3}\Phi''(x_c)\mathcal{H}_{ppp}(x_c,0)}{\mathcal{H}_{pp}(x_c,0)}.\tag{D.3}$$

Note that $H(x,p)\neq\mathcal{H}(x,p)$, where $\mathcal{H}(x,p)$ is the Hamiltonian (4.29).

E $x\to 0$ limit of the quasi-stationary density

The WKB approximation (4.49) of the discrete process brakes down in the limit $x\to 0$, due to small copy number effects (i.e., fluctuations are on the same order). This fact is not relevant if one is interested only in approximating the mean exit time. However, we also approximate the effective potential. Although $\Phi(x)$ is bounded in the limit $x\to 0$, $\Psi(x)$ has a logarithmic singularity. To correct this, we use the discrete master equation (2.4) to calculate $\boldsymbol{\phi}_0(0)$, with

$$\boldsymbol{\phi}_0(0)=-(\alpha_{\rm i}A(0)-\alpha_{\rm e}\Sigma_{\mathbf{v}(0)})^{-1}\boldsymbol{\phi}_0(\frac{1}{\alpha_{\rm e}}).\tag{E.1}$$

The WKB approximation can be use for the value of the eigenfunction near the boundary, with

$$\boldsymbol{\phi}_0(1/\alpha_\mathrm{e}) = \hat{\mathbf{w}}(1/\alpha_\mathrm{e})k(1/\alpha_\mathrm{e}) \exp\left[-\frac{1}{\epsilon}\Phi(1/\alpha_\mathrm{e})\right]. \tag{E.2}$$

Hence,

$$\boldsymbol{\phi}_0(0) = -(\alpha_\mathrm{i}A(0) - \alpha_\mathrm{e}\Sigma_{\mathbf{v}(0)})^{-1}\hat{\mathbf{w}}(1/\alpha_\mathrm{e})k(1/\alpha_\mathrm{e}) \exp\left[-\frac{1}{\epsilon}\Phi(1/\alpha_\mathrm{e})\right]. \tag{E.3}$$

References

[1] Michael Assaf, Elijah Roberts, and Zaida Luthey-Schulten. Determining the stability of genetic switches: Explicitly accounting for mrna noise. *Phys. Rev. Lett.*, 106(24):248102, Jun 2011.

[2] D. J. Bicout. Green's functions and first passage time distributions for dynamic instability of microtubules. *Phys. Rev. E*, 56:6656–6667, 1997.

[3] Paul C. Bressloff. Metastable states and quasicycles in a stochastic wilson-cowan model of neuronal population dynamics. *Phys. Rev. E*, 82(5):051903, Nov 2010.

[4] Charles Doering, Khachik Sargsyan, and Leonard Sander. Extinction times for birth-death processes: Exact results, continuum asymptotics, and the failure of the fokker–planck approximation. *Multiscale Model. Simul.*, 3(2):283–299, 2005.

[5] Charles R Doering, Khachik V Sargsyan, Leonard M Sander, and Eric Vanden-Eijnden. Asymptotics of rare events in birth–death processes bypassing the exact solutions. *Journal of Physics: Condensed Matter*, 19(6):065145, 2007.

[6] M. I. Dykman, Eugenia Mori, John Ross, and P. M. Hunt. Large fluctuations and optimal paths in chemical kinetics. *J. Chem. Phys.*, 100(8):5735–5750, 1994.

[7] Carlos Escudero and Alex Kamenev. Switching rates of multistep reactions. *Phys. Rev. E*, 79:041149, Apr 2009.

[8] Jin Feng and Thomas G Kurtz. *Large deviations for stochastic processes*, volume v. 131 of *Mathematical surveys and monographs*. American Mathematical Society, Providence, R.I., 2006.

[9] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer- Verlag, New York, 2nd edition edition, 1998.

[10] A. Friedman and G. Craciun. A model of intracellular transport of particles in an axon. *J. Math. Biol.*, 51(2):217–246, Aug 2005.

[11] C. W Gardiner. *Handbook of stochastic methods for physics, chemistry, and the natural sciences*, volume v. 13. Springer-Verlag, Berlin, 1983.

[12] Peter Hanggi, Hermann Grabert, Peter Talkner, and Harry Thomas. Bistable systems: Master equation versus fokker-planck modeling. *Phys. Rev. A*, 29(1):371–378, Jan 1984.

[13] Matthias Heymann and Eric Vanden-Eijnden. The geometric minimum action method: A least action principle on the space of curves. *Communications on Pure and Applied Mathematics*, 61(8):1052–1117, 2008.

[14] James M. Hill and Barry D. Hughes. On the general random walk formulation for diffusion in media with diffusivities. *The ANZIAM Journal*, 27:73–87, 6 1985.

[15] R Hinch and S. J. Chapman. Exponentially slow transitions on a markov chain: the frequency of calcium sparks. *Eur. J. Appl. Math.*, 16(Part 4):427–446, Aug 2005.

[16] James P. Keener and Jay M. Newby. Perturbation analysis of spontaneous action potential initiation by stochastic ion channels. *Phys. Rev. E*, 84(1):011918, 2011.

[17] T. B. Kepler and Timothy C. Elston. Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys. J.*, 81(6):3116–3136, 2001.

[18] Andras Kramli and Domokos Szasz. Random walks with internal degrees of freedom. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 63:85–95, 1983.

[19] Yueheng Lan, Timothy C Elston, and Garegin A Papoian. Elimination of fast variables in chemical langevin equations. *J. Chem. Phys.*, 129(21):214115, Dec 2008.

[20] Uzi Landman, Elliott W. Montroll, and Michael F. Shlesinger. Random walks and generalized master equations with internal degrees of freedom. *Proceedings of the National Academy of Sciences*, 74(2):430–433, 1977.

[21] J. Y. Lee and M. J. Ward. On the asymptotic and numerical-analyses of exponentially ill-conditioned singularly perturbed boundary-value-problems. *Stud. Appl. Math.*, 94(3):271–326, APR 1995.

[22] Donald Ludwig. Persistence of dynamical systems under random perturbations. *SIAM Review*, 17(4):pp. 605–640, 1975.

[23] Robert S. Maier and Daniel L. Stein. Limiting exit location distributions in the stochastic exit problem. *SIAM J. Appl. Math.*, 57(3):752–790, 1997.

[24] B. J. Matkowsky, Z. Schuss, and C. Tier. Diffusion across characteristic boundaries with critical points. *SIAM J. Appl. Math.*, 43(4):673–695, 1983.

[25] Philipp Metzner, Christof Schütte, and Eric Vanden-Eijnden. Transition path theory for markov jump processes. *Multiscale Model. Simul.*, 7(3):1192–1219, 2009.

[26] T. Naeh, M. M. Klosek, B. J. Matkowsky, and Z. Schuss. A direct approach to the exit problem. *SIAM J. Appl. Math.*, 50(2):pp. 595–627, 1990.

[27] Jay Newby and Paul C Bressloff. Local synaptic signaling enhances the stochastic transport of motor-driven cargo in neurons. *Physical Biol.*, 7(3):036004, 2010.

[28] Jay M Newby. Isolating intrinsic noise sources in a stochastic genetic switch. *Physical Biology*, 9(2):026002, 2012.

[29] Jay M. Newby and Paul C. Bressloff. Directed intermittent search for a hidden target on a dendritic tree. *Phys. Rev. E*, 80(2, Part 1):021913, 2009.

[30] Jay M. Newby and James P. Keener. An asymptotic analysis of the spatially inhomogeneous velocity-jump process. *Multiscale Model. Simul.*, 9(2):735–765, 2011.

[31] HG Othmer, SR Dunbar, and W Alt. Models of dispersal in biological-systems. *J. Math. Biol.*, 26(3):263–298, 1988.

[32] Zeev Schuss. *Theory and applications of stochastic processes: an analytical approach*, volume v. 170 of *Applied mathematical sciences*. Springer, New York, 2010.

[33] Adam Shwartz and Alan Weiss. *Large deviations for performance analysis: queues, communications, and computing*. Stochastic modeling series. Chapman & Hall, London, 1st ed edition, 1995.

[34] P. Talkner. Mean first passage time and the lifetime of a metastable state. *Zeitschrift für Physik B Condensed Matter*, 68:201–207, 1987.

[35] Mukund Thattai and Alexander van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.*, 98(15):8614–8619, 2001.

[36] N. G. Van Kampen. Composite stochastic-processes. *Physica A*, 96(3):435–453, 1979.

[37] Melissa Vellela and Hong Qian. A quasistationary analysis of a stochastic chemical reaction: Keizer's paradox. *Bull. Math. Biol.*, 69:1727–1746, 2007.

[38] Aleksandra M. Walczak, José N. Onuchic, and Peter G. Wolynes. Absolute rate theories of epigenetic stability. *Proceedings of the National Academy of Sciences of the United States of America*, 102(52):18926–18931, 2005.